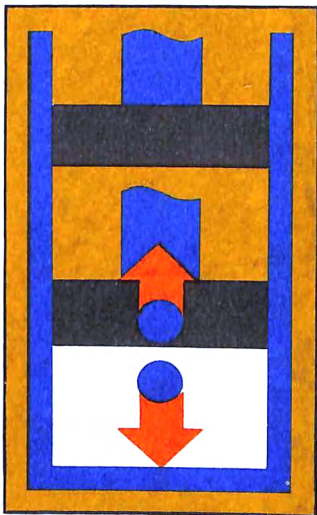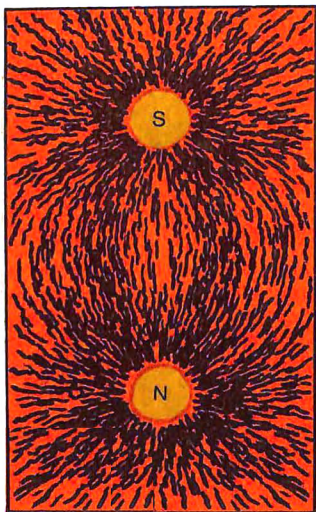Thermal Phenomena · Molecular Physics
Fundamentals of Electrodynamics

Senior

PHYSICS 2

B.B.Bukhovtsev,Yu.L.Klimontovich,G.Ya.Myakishev
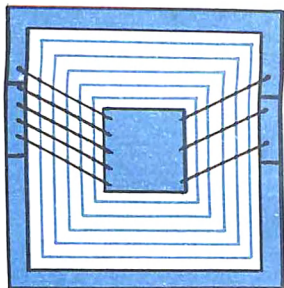
Substance	Specific heat of vaporization, kJ/kg
Water	2260
Ethyl alcohol	860
Nitric acid	480
Ether	360
Mercury	290
Kerosene	210

Substance	Critical temperature, °C
Helium	-268
Hydrogen	-240
Nitrogen	-147
Oxygen	-118
Chlorine	146
Ether	194
Mercury	1460

Material	Modulus of elasticity, GPa (at $t = 20\ °C$)
Aluminium	70-71
Concrete	14.3–23.2
Laminated plastic	10–18
Iron	190–210
Gold	79
Rubber	0.9
Steel	200–220
Titanium	112

Temperature, °C	Saturated vapour pressure for water, kPa
0	0.421
10	1.22
15	1.71
20	2.33
25	3.17
30	4.27
50	12.3
80	46.7
100	101.3

Senior

PHYSICS 2

B.B.Bukhovtsev, Yu.L.Klimontovich, G.Ya.Myakishev

Mir Publishers Moscow

Contents

6

7

Thermal Phenomena
Molecular Physics

Introduction

MECHANICAL MOTION. The mechanical motion of matter, i.e. the displacement of some bodies in space relative to other bodies with time, was considered in detail in "Senior Physics 1". The fact that all bodies consist of atoms and molecules was then disregarded. Objects were treated as continuous bodies devoid of internal structure.

Mechanics does not aim at the investigation of the properties of bodies. It studies the position of bodies in space and their velocities at any instant of time.

THERMAL MOTION. As is well known from the course of "Junior Physics", atoms and molecules of any substance are in random (chaotic) motion called the THERMAL MOTION. In the first part of the present book, we shall describe the basic laws governing the thermal motion of matter.

Molecular motion is chaotic because the number of molecules in bodies is unimaginably large. Every molecule incessantly changes its velocity upon collisions with other molecules. As a result, its trajectory is extremely complicated, and its motion is random. The concept of thermal motion is inapplicable to systems containing a few molecules. Random motion of a huge number of molecules differs in principle from ordered mechanical motion of individual bodies. This is why thermal motion represents a peculiar form of motion of matter, which has specific properties.

Thermal motion is responsible for the internal properties of bodies. By analyzing this type of motion, we can explain many physical processes occurring in bodies.

MACROSCOPIC BODIES. Bodies consisting of a very large number of atoms or molecules are called MACROSCOPIC BODIES. The size of macroscopic bodies exceeds the atomic size by several orders of magnitude. Gas in a cylinder, water in a glass, a grain of sand, a stone, a steel rod and a globe are all examples of macroscopic bodies (Fig. 1).

We shall consider processes in macroscopic bodies.

THERMAL PHENOMENA. Thermal motion of molecules depends on the temperature. This fact was pointed out in the courses of "Junior
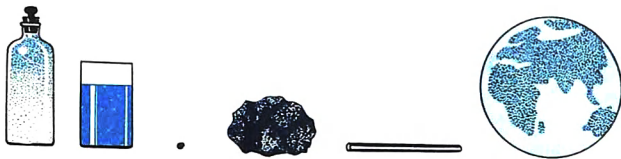
Fig. 1

Physics". By analyzing the thermal motion of molecules, we shall study the phenomena that depend on the temperature of bodies. Upon heating, substances may go over from one state of aggregation to another: solid bodies are liquefied, and liquids evaporate. On the contrary, gases when cooled are converted into liquids, while liquids become solids.

These and many other phenomena caused by random motion of atoms and molecules are known as THERMAL PHENOMENA.

THE ROLE OF THERMAL PHENOMENA. Thermal phenomena play a significant role in the life of human beings, animals and plants. A change in the air temperature by 20-30 °C with a change of seasons in Europe alters the surroundings. When spring comes, nature wakes up, woods and meadows turn green. In winter the rich hues of summer are replaced by a monotonous white, and the life of plants and many insects comes to a standstill. When the temperature of our body rises just by a degree, we feel feverish.

Thermal phenomena have attracted the attention of man since ancient time. Mankind gained a relative independence from environmental conditions when it learned to strike a fire and keep it burning. This was one of the most important discoveries made by man.

A change in temperature affects all the properties of bodies. For example, heating or cooling changes the shape of solids and the volume of liquids. Their mechanical properties, such as elasticity, also vary considerably. A piece of a rubber tube does not change much after having been struck by a hammer. However, when cooled to below − 100 °C, rubber becomes as brittle as glass. A slight impact breaks it into small pieces. Only after heating does rubber regain its elastic properties.

The thermal phenomena listed above, as well as many other phenomena, are governed by certain laws. The discovery of these laws makes it possible to apply thermal phenomena in engineering. Modern heat engines, apparatus for liquefaction of gases, refrigerators and other machines are designed on the basis of these laws.

MOLECULAR KINETIC THEORY. The molecular kinetic theory explains thermal phenomena in macroscopic bodies and their internal properties proceeding from the assumption that *all bodies consist of individual particles which are in random motion*. This theory tries to associate the regularities in the behaviour of individual molecules with the quantities characterizing the properties of macroscopic bodies.

Even ancient philosophers guessed that heat is a form of internal motion

Mikhail Vasil'evich Lomonosov (1711–1765) was a great Russian scientist with an encyclopaedic knowledge, poet and social worker, and the founder of Moscow University named after him. Pushkin called Lomonosov "the first Russian university". Lomonosov was the author of outstanding works on physics, chemistry, mining and metallurgy. He developed the molecular kinetic theory of heat and predicted the laws of mass and energy conservation. Lomonosov created fundamental works on the history of the Russian people. He was the founder of the modern Russian grammar.

of particles constituting a body. An important contribution to the evolution of the molecular kinetic theory was made by the great Russian scientist M. V. Lomonosov. He considered heat as the rotational motion of the particles constituting a substance. Using this theory, he successfully explained the phenomena of fusion, evaporation and heat conduction. He came to the conclusion about the existence of the "extreme, or ultimate degree of cold", when the motion of particles ceases.

The difficulties in the construction of the molecular kinetic theory, however, delayed its triumph which was achieved only at the beginning of the 20th century. As a matter of fact, the number of molecules in macroscopic bodies is so large that it is impossible to trace the motion of an individual molecule. It is necessary to learn to apply the laws of motion of individual molecules to finding the average result of their combined motion. It is this average result that determines the thermal phenomena in macroscopic bodies.

THERMODYNAMICS. A substance possesses many properties that can be studied without a thorough knowledge of its structure. Thermal phenomena can be described with the help of quantities registered by such instruments as a manometer or a thermometer which do not respond to the action of individual molecules.

The first scientific theory of thermal processes – THERMODYNAMICS – was constructed in the middle of the 19th century, following the discovery of the law of energy conservation. Thermodynamics is the theory of thermal phenomena which does not take into account the molecular structure of bodies. It was formulated while studying the optimal conditions for using heat to gain work long before the molecular kinetic theory received world-wide recognition.

THERMODYNAMICS AND STATISTICAL MECHANICS. In the modern science and engineering, both thermodynamics and the molecular kinetic theory, also called STATISTICAL MECHANICS, are used. These theories complement each other.

The entire subject of thermodynamics is contained in a few statements called the LAWS OF THERMODYNAMICS. These laws were established experimentally. They are applicable to all substances irrespective of their internal structure. Statistical mechanics is a deeper and a more complicated theory of thermal phenomena. This theory can be used to substantiate theoretically all the laws of thermodynamics.

We shall begin with the basic concepts of the molecular kinetic theory. partially described in the courses of "Junior Physics". After this. we shall discuss the quantitative theory of a simple system, viz. a low-density gas.

1
FUNDAMENTALS OF MOLECULAR KINETIC THEORY

1.1. Basic Concepts. Dimensions of Molecules

The molecular kinetic theory of matter is based on three statements, each of which has been proved experimentally: (1) a substance consists of particles, (2) these particles move at random, and (3) the particles interact.

The properties and behaviour of all bodies are determined by the motion of interacting particles constituting them, viz. molecules, atoms[1] or still smaller particles known as elementary particles.

ESTIMATION OF MOLECULAR DIMENSIONS. To make sure that molecules actually exist their dimensions should be determined.

Let us consider a relatively simple method of estimating molecular dimensions. It is known that a drop of olive oil having a volume of 1 mm³ cannot be spread to cover more than 0.6 m² of the surface of water. It can be assumed that when the drop of oil spreads to cover the maximum area, the layer thickness is just equal to the diameter of a molecule. The thickness of this layer can easily be determined, and thus the dimensions of a molecule of olive oil can be estimated.

The volume V of the layer of oil is equal to the product of its area S by the thickness d: $V = Sd$ (Fig. 2). Consequently, the thickness of the layer, and thus the size of the olive oil molecule, is given by

$$d = \frac{0.001 \ \text{cm}^3}{6000 \ \text{cm}^2} \simeq 1.7 \times 10^{-7} \ \text{cm}.$$

ION PROJECTOR. There is no need to list all possible methods available at present to prove the existence of atoms and molecules. Modern instruments, such as electron microscope, make it possible to observe the

---

[1] It should be recalled that the atom is the smallest particle of a chemical element possessing its properties, and the molecule is the smallest stable particle of a substance, having its properties. Molecules are formed by atoms.

Fig. 2
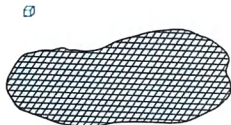
images of individual atoms and molecules. However, the electron microscope is a very complex instrument. We shall consider a much simpler device which makes it possible to obtain images of individual atoms and estimate their size. This is the ION PROJECTOR, or ion microscope. It consists of a spherical vessel of about 10 cm in radius containing a tungsten needle whose point is arranged at the centre of the vessel (Fig. 3). The radius of curvature of the needle point is made as small as it is possible with modern methods of metal processing (viz. about 5×10^{-6} cm). The inner surface of the sphere is coated by a thin conducting layer which, like the screen of a kinescope, can glow under the impacts of fast particles. A voltage of a few hundred volts is applied between the negatively charged conducting layer and the positively charged needle point. The vessel is filled with helium under a low pressure of 100 Pa (0.75 mm Hg).

Tungsten atoms form microscopic "tubercles" on the surface of the needle point (Fig. 4). As randomly moving helium atoms come close to tungsten atoms, the electric field, which is especially strong at the needle point, tears out electrons from helium atoms and converts them into positive ions. Helium ions are repelled by the positively charged needle point and move at high velocities along the radii of the sphere. Under the impacts of helium ions, the surface of the sphere starts to glow. Figure 5 shows the arrangement of atoms on the point of a tungsten needle. Light spots are the images of individual atoms.

The dimensions of molecules, including those of olive oil, are larger than atomic dimensions. Atomic diameters are of the order of 10^{-8} cm. This size is so small that it is impossible to vizualise it. In such a case we resort to
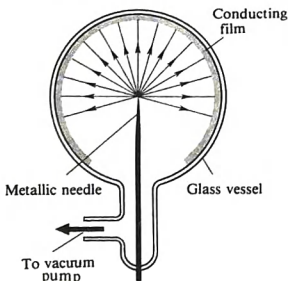
Fig. 3
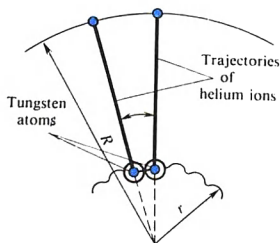
Fig. 4

14
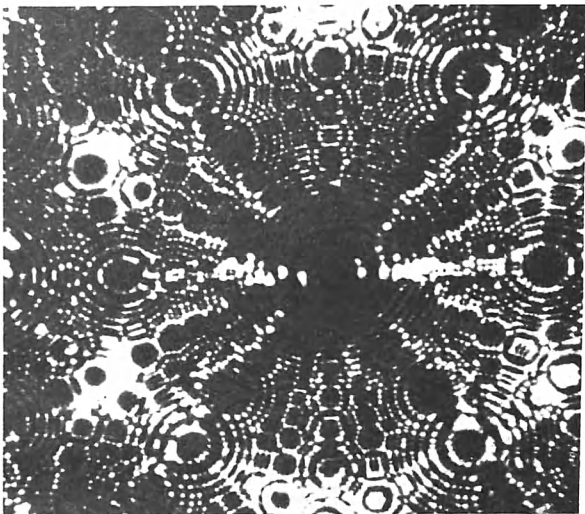
Fig. 5

comparisons. One of them can be as follows. If we magnify a fist to the size of the Earth, an atom will be of the size of the fist upon the same magnification.

NUMBER OF MOLECULES. Since molecules are very small, the number of molecules in any macroscopic body is extremely large. Let us calculate approximately the number of molecules in a water drop having a mass of 1 g, and hence a volume of 1 cm³. The diameter of a water molecule is approximately equal to 3×10^{-8} cm. Assuming that each water molecule occupies a volume of $(3 \times 10^{-8}$ cm$)^3$, we can find the number of molecules in the closely packed drop by dividing its volume (1 cm³) by the volume of a molecule:

$$N = \frac{1}{(3 \times 10^{-8})^3} \simeq 3.7 \times 10^{22}.$$

When you take a breath, you get into your lungs so many molecules that if they were uniformly distributed in the atmosphere of the Earth upon exhalation, every inhabitant of our planet would inhale two of the molecules that had been in your lungs.

1.2. Molecular Mass. Avogadro's Number

THE MASS OF A WATER MOLECULE. The masses of individual molecules and atoms are very small. For example, 1 g of water contains 3.7×10^{22} molecules. Consequently, the mass of a water molecule is

$$m_{0H_2O} \simeq \frac{1}{3.7 \times 10^{22}} \text{ g} \simeq 3 \times 10^{-23} \text{ g}. \qquad (1.1)$$

The masses of molecules of other substances are of the same order of magnitude (excluding huge molecules of organic compounds).

RELATIVE MOLECULAR MASS. Since the masses of molecules are very small, it is convenient to use in calculations the relative values of masses instead of their absolute values. According to an international agreement, the masses of all atoms and molecules are compared with 1/12 of the mass of a carbon atom (the so-called carbon scale of atomic masses)[1]. The relative molecular (or atomic) mass M_r of a substance is the ratio of the mass m_0 of a molecule (or atom) of this substance to 1/12 of the mass m_{0C} of a carbon atom:

$$M_r = \frac{m_0}{(1/12) m_{0C}}. \qquad (1.2)$$

The relative atomic masses of all chemical elements are known at present to a high degree of accuracy. Adding the relative atomic masses of elements constituting the molecules of a substance, we can calculate the relative molecular mass. For example, the relative molecular mass of carbon dioxide CO_2 is approximately equal to 44 since the relative atomic mass of carbon is 12 and that of oxygen is nearly 16: $12 + 2 \times 16 = 44$.

AMOUNT OF SUBSTANCE AND AVOGADRO'S NUMBER. It would be most natural to measure the amount of a substance in terms of the number of molecules or atoms in the body. But since the number of molecules in any macroscopic body is very large, the relative number of molecules is used in calculations instead of the absolute number.

In the International System of Units (SI), the amount of substance is expressed in MOLES. A mole is the amount of substance containing the same number of molecules or atoms as the number of atoms in 0.012 kg of carbon.

Naturally, one mole of any substance contains the same number of atoms or molecules. This number is denoted by N_A and is known as AVOGADRO'S NUMBER after the famous Italian scientist of the 19th century.

---

[1] A comparison of atomic and molecular masses with 1/12 of the mass of the carbon atom was adopted in 1961. The main reason behind such a choice is that carbon is a constituent of a very large number of chemical compounds. The factor 1/12 was introduced to make the relative masses of atoms closer to integers. The relative mass of carbon is strictly equal to 12, while that of hydrogen is close to unity.

In order to determine Avogadro's number, it is necessary to calculate the mass of a carbon (or any other) atom. A rough estimate can be obtained in the same way as it was done above for a water molecule (more accurate methods are based on the deflection of ion beams by an electromagnetic field). Measurements give the following value for the atomic mass of carbon: $m_{0C} = 1.995 \times 10^{-26}$ kg.

Avogadro's number N_A can be determined by dividing the mass of a mole of carbon by the mass of a carbon atom:

$$N_A = 0.012 \frac{\text{kg}}{\text{mole}} \frac{1}{1.995 \times 10^{-26} \text{ kg}} = 6.02 \times 10^{23} \text{ mole}^{-1}.$$

The notation mole^{-1} indicates that N_A is the number of atoms of any substance in a mole of this substance. If the amount of substance is $\nu = 2.5$ mole, the number of molecules in this substance is $N = \nu N_A \simeq 1.5 \times 10^{24}$. Thus, the amount of substance ν is equal to the ratio of the number N of molecules in a given body to Avogadro's number N_A, i.e. to the number of molecules in a mole of the substance:

$$\nu = \frac{N}{N_A}. \tag{1.3}$$

The large value of Avogadro's number reflects the smallness of microscopic scales in comparison with macroscopic ones. A body containing a mole of substance has conventional macroscopic dimensions.

MOLAR MASS. Along with the relative mass M_r, the molar mass M is also widely used in physics and chemistry. The molar mass is the mass of a mole of a substance.

According to this definition, the molar mass is equal to the product of the mass of a molecule and Avogadro's number:

$$M = m_0 N_A. \tag{1.4}$$

The mass m of any amount of substance is equal to the product of the molecular mass and the number of molecules in the body:

$$m = m_0 N. \tag{1.5}$$

Replacing in formula (1.3) N_A and N by their expressions from (1.4) and (1.5), we obtain

$$\nu = \frac{m}{M}. \tag{1.6}$$

The amount of substance is equal to the ratio of the mass of the substance to its molar mass.

According to formulas (1.3) and (1.6), the number of molecules of a substance having a mass m and a molar mass M is

$$N = \nu N_A = N_A \frac{m}{M}. \tag{1.7}$$

1. List the basic statements of the molecular kinetic theory.
2. Give proofs of the existence of molecules.
3. Give the definition of the relative molecular mass.
4. What is the relative molecular mass of water?
5. What do we call the amount of substance?
6. Define the unit of the amount of substance, viz. a mole.
7. What is the value of Avogadro's number?
8. Give the definition of the molar mass.
9. What is the relation between the mass of a body and the amount of substance in it?
10. Prove that the molar mass M is connected with the relative molecular mass through the following relation:
    $M = 10^{-3} \cdot M_r$, $kg \cdot mole^{-1}$.
11. What is the value of the molar mass for carbon dioxide CO_2?
12. What is the mass of 3 moles of carbon dioxide?

1.3. Brownian Movement

In "Junior Physics", we considered the phenomenon of diffusion, viz. the mixing of gases, liquids and solids in direct contact. This phenomenon can be explained by the random motion of molecules. But the most convincing proof of the molecular motion can be obtained by observing through a microscope the tiny particles of a solid suspended in water. These particles are in random motion which is known as the BROWNIAN MOVEMENT.

The Brownian movement is the thermal motion of particles suspended in a liquid (or gas).

OBSERVATION OF BROWNIAN MOVEMENT. The English botanist Brown observed this phenomenon for the first time in 1827 while examining clubmoss spores suspended in water through a microscope. In modern experiments on Brownian movement, the particles of gamboge gum dye, which are insoluble in water, are normally used. These particles are in random motion. The most striking and unusual fact is that this movement never ceases. We are accustomed that any moving body ultimately stops. However, *Brownian movement is the thermal motion which cannot be terminated.* Its intensity increases with temperature. Figure 6 is a schematic diagram of Brownian movement. The positions of the particles marked by dots are determined in equal time intervals of 30 s. These points are connected by segments of straight lines. Actually, the trajectories of particles are much more complicated.

Brownian movement can also be observed in a gas. Dust and smoke particles suspended in air are in Brownian movement.

Nowadays, the concept of Brownian movement is used in a wider sense. For example, the oscillation of pointers in high-precision measuring instruments is a kind of Brownian movement. It occurs due to the thermal motion of atoms constituting the parts of instruments and the ambient.

EXPLANATION OF BROWNIAN MOVEMENT. Brownian movement can be explained only with the help of the molecular kinetic theory. The
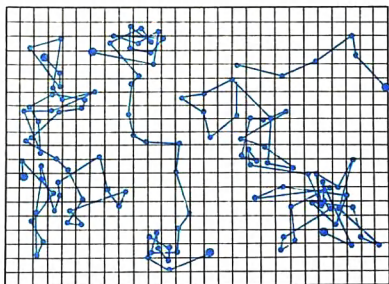
Fig. 7

Fig. 6

reason behind the Brownian movement of a particle is that the impacts of molecules against this particle are not balanced. Figure 7 shows schematically the position of a Brownian particle and its neighbouring molecules. In random motion of molecules, the momenta transferred to the Brownian particle, say, from left and right, are not equal. Therefore, the resultant force of pressure, that causes a change in the motion of the Brownian particle, differs from zero.

The average pressure has a certain value in a gas as well as in a liquid. However, insignificant random deviations from the average value always take place. The smaller the surface area of a body, the larger the relative variations of the force of pressure acting on this surface. Thus, if the surface area is of the order of several molecular diameters, the force of pressure acting on this area varies abruptly from zero to a certain value when this area is hit by a molecule.

The molecular kinetic theory of Brownian movement was constructed by A. Einstein in 1905. The formulation of the theory of Brownian movement and its experimental verification by the French physicist J. Perrin crowned the victory of the molecular kinetic theory.

1.4. Forces of Intermolecular Interaction

It is not difficult to demonstrate the existence of considerable forces of interaction among atoms and molecules. Just try to break a thick rod! This rod consists of molecules. If there were no attractive forces among molecules, all substances would be in the gaseous state under any conditions. It is due to attractive forces that molecules are kept close to one another and form liquids and solids. However, attractive forces alone cannot ensure the existence of stable complexes of atoms and molecules. At very short distances, molecules repel one another. For this reason, molecules do not penetrate into each other.

19

STRUCTURE OF MOLECULES. A molecule is a complex system consisting of individual charged particles, viz. electrons and atomic nuclei. Although molecules are electrically neutral as a whole, considerable electric forces act between them at short distances. Interaction takes place between electrons and nuclei of neighbouring molecules. The description of the motion of particles within atoms and molecules and the FORCES OF INTERMOLECULAR INTERACTION is a very complicated problem. This problem is considered in atomic physics. Here we shall give only the result of its solution, i.e. an approximate dependence of the force of interaction of two molecules on their separation.

Molecules consist of particles bearing unlike charges. Attractive forces act between the electrons of one molecule and the atomic nuclei of another molecule. At the same time, repulsive forces act between the electrons of the two molecules and between their nuclei. The resultant of the forces of intermolecular interaction is equal to the difference in the magnitudes of these forces.

DEPENDENCE OF MOLECULAR FORCES ON INTERMOLECULAR DISTANCES. Let us now analyze how the projection of the force of interaction between two molecules onto the straight line connecting them depends on the distance between the molecules.

If the molecules are at a distance exceeding several times their dimensions, the forces of interaction are practically absent. The forces acting between electrically neutral molecules are of the SHORT-RANGE NATURE. The repulsive force is zero at distances two or three times longer than the molecular diameter. Only the attractive force is noticeable at such distances. As the intermolecular distance decreases, the attractive force grows at first. However, with a further decrease in the separation between the molecules, the repulsive force becomes significant. This force increases very rapidly when the electron shells of the atoms begin to overlap.

Figure 8 represents the approximate dependence of the projections of the repulsive force (upper curve), of the attractive force (lower curve) and of the resultant force (middle curve) on the centre-to-centre distance between molecules. The projection of the repulsive force is positive, while the projection of the attractive force is negative. Thin vertical lines are drawn to make the composition of the projections of forces more convenient.

At a distance r_0, approximately equal to the sum of the molecular radii, the projection of the resultant force F_r is equal to zero since the attractive force is equal in magnitude to the repulsive force (Fig. 9a). For $r > r_0$, the attractive force exceeds the repulsive force, and the projection of the resultant force (bold arrow) is negative (Fig. 9b). As $r \to \infty$, $F_r \to 0$. At distances $r < r_0$, the repulsive force is stronger than the attractive force (Fig. 9c).

THE ORIGIN OF ELASTIC FORCES. The dependence of the forces of intermolecular interaction on the separation of molecules explains the emergence of elastic force during compression or extension of bodies. If we try to bring molecules to a distance shorter than r_0, the force preventing the molecules from coming closer begins to act. On the contrary, when the molecules are moved apart, the attractive force emerges, returning the

Fig. 8

Forces acting on the right molecule
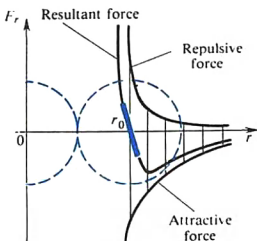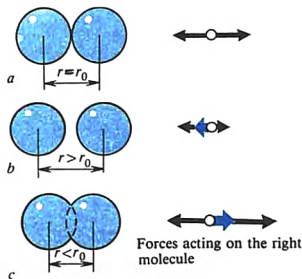
Fig. 9

molecules to their initial positions after the external force is removed.

When the displacements of the molecules from their equilibrium position are small, attractive or repulsive forces change linearly with the displacement. On a short segment, the curve can be treated as a straight line (the middle part of the curve in Fig. 8). Therefore, Hooke's law is valid for small deformations. It states that the elastic force is proportional to the deformation. For large displacements of molecules, Hooke's law is no longer applicable.

Since the distances between all molecules of a body change upon its deformation, the adjoining molecular layers are displaced only to a very small extent. For this reason, Hooke's law is valid for deformations exceeding the molecular dimensions by millions of times.

1.5. Structure of Gases, Liquids and Solids

The molecular kinetic theory makes it possible to explain why a substance can be in the gaseous, liquid or solid state.

GASES. The distances between atoms and molecules in gases are on the average many times larger than the dimensions of the molecules themselves (Fig. 10). For example, under atmospheric pressure, the volume of a vessel exceeds the volume of gas molecules in it by about four orders of magnitude.

Gases can easily be compressed since only the average intermolecular distance decreases upon the compression, but the molecules do not "squeeze" one another (Fig. 11).

Gas molecules move in space at enormous velocities (of the order of hundreds of meters per second). Colliding with one another, they bounce like billiard balls in different directions.

Weak attractive forces between gas molecules are unable to keep them close. For this reason, gases can expand without limits. They retain neither shape nor volume.

21
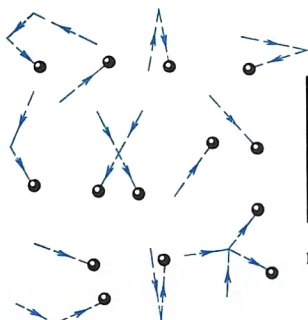
Fig. 10
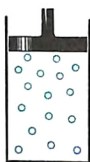
Fig. 11

Fig. 12

Numerous impacts of gas molecules against vessel walls are responsible for the gas pressure.

LIQUIDS. Molecules in liquids are arranged very close to one another (Fig. 12), and therefore liquid molecules behave not like gas molecules. Being squeezed by neighbouring molecules as in a cage, a molecule in a liquid as if "marks time" (vibrates about its equilibrium position upon collisions with neighbouring molecules). It only "jumps" from time to time, breaking through the "bars of the cage", but immediately gets into another "cage" formed by new neighbours. The time of the "sedentary life" of a water molecule, i.e. the time during which it vibrates about a certain equilibrium position, is on the average about 10^{-11} s at room temperature. The time of a single vibration is much shorter (10^{-12}-10^{-13} s). As the temperature increases, the "sedentary lifetime" of molecules decreases. The nature of molecular motion in liquids, which was first established by the Soviet physicist Ya. Frenkel, allows us to explain the basic properties of liquids.

Molecules in a liquid are packed very closely. Therefore, if we try to decrease the volume of the liquid even slightly, the molecules themselves are deformed (Fig. 13). But to deform molecules, large forces are required. This explains the poor compressibility of liquids.

Fig. 13

Fig. 14

Fig. 15

Fig. 16

22
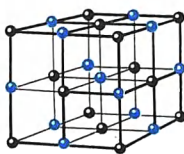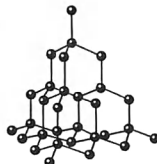
Yakov Il'ich Frenkel (1894–1952) was an outstanding Soviet theoretical physicist who made a considerable contribution to various branches of physics. Frenkel was the author of the modern theory of the liquid state of matter. He laid the foundation of the theory of ferromagnetism. His works on atmospheric electricity and the origin of the magnetic field of the Earth have received worldwide recognition. The first quantitative theory for the fission of uranium nuclei was developed by Frenkel.

It is well known that liquids exhibit fluidity, i.e. they do not retain their shape. This can be explained as follows. If a liquid is at rest, the molecules jump from one "sedentary" position to another at the same frequency in all directions (see Fig. 12). When an external force is applied, it does not noticeably change the number of jumps per second. However, in this case the molecules jump from one "sedentary" position to another predominantly in the direction of the external force (Fig. 14). For this reason, liquids flow and assume the shape of the vessel.

SOLIDS. Unlike liquid molecules, atoms and molecules in solids vibrate about definite equilibrium positions. True, solid molecules sometimes alter their equilibrium position, but this occurs quite seldom. For this reason, solid bodies retain their shape and volume.

Fig. 17

23

There is one more important difference between liquids and solids. A liquid can be compared with a crowd whose individual members are jostling restlessly on the spot, while a solid resembles a formed squadron in which although the soldiers are not at rest (due to the thermal motion), but keep on the average strict intervals between the members. If we connect the equilibrium positions of atoms or ions in a solid, we obtain an orderly spatial network called the CRYSTAL LATTICE. Figures 15 and 16 represent crystal lattices of common salt and diamond. The intrinsic order in the arrangement of atoms in a crystal leads to a geometrically regular external shape. Figure 17 shows diamonds found in the Yakutsk deposit.

?

1. Give proofs of the existence of thermal motion of molecules.
2. Why is Brownian movement noticeable only for light particles?
3. What is the nature of intermolecular forces?
4. What is the relation between the intermolecular forces and the separations between the molecules?
5. Two lead bricks with smooth clean cuts are brought in contact and stick together. Explain this phenomenon.
6. What is the difference in the thermal motion of molecules in solids, liquids and gases?

1.6. Ideal Gas in Molecular Kinetic Theory

The qualitative explanation of the basic properties of substances on the basis of the molecular kinetic theory is not very complicated. However, the theory that establishes quantitative relations between the experimentally measured values (pressure, temperature, and so on) and the properties of molecules themselves, their number and the velocity of motion is quite cumbersome. We shall confine ourselves to the theory of sufficiently rarefied gases.

IDEAL GAS. Intermolecular distances in a rarefied gas greatly exceed the dimensions of molecules. In this case, the intermolecular interactions are negligibly small, and the kinetic energy of molecules is many times higher than the potential energy of interaction. Gas molecules can be treated as very small rigid balls. There is no attraction between the molecules (balls), while the repulsive forces are manifested only over very short time intervals during molecular collisions.

Instead of a REAL GAS with complex forces of interaction between the molecules, we shall consider its PHYSICAL MODEL. This model is known as an IDEAL GAS. An ideal gas is a gas in which the intermolecular interaction is negligibly small.

In a physical model, only those properties of a real system are considered that have to be taken into account for explaining the regularities in the behaviour of the system to be studied.

None of the models can reflect ALL the properties of a system. However, we shall consider here a sufficiently narrow problem. Using the molecular kinetic theory, we shall calculate the pressure of a rarefied gas on the vessel
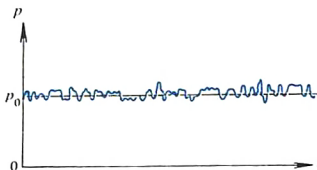
Fig. 18

walls. The ideal gas model turns out to be quite satisfactory for this problem. It leads to the results which are confirmed by experiment. REAL RAREFIED GASES INDEED BEHAVE LIKE AN IDEAL GAS.

GAS PRESSURE IN THE MOLECULAR KINETIC THEORY. Suppose that a gas is contained in a closed vessel. A manometer indicates that its pressure is p_0[1]. But how does this pressure emerge? Every gas molecule impinging at the vessel wall acts on it with a certain force during a short interval of time. As a result of random impacts against the wall, the force exerted by all the molecules per unit area of the wall, viz. the pressure, rapidly varies with time as shown in Fig. 18. However, the effects due to the impacts of individual molecules are so weak that they cannot be registered by the manometer. The manometer indicates the time-averaged force acting per unit area of its sensitive element, viz. the membrane. Despite small variations of pressure, its average value p_0 is practically found to be constant, since the number of molecular collisions is very large, while the mass of the molecules is very small.

The gas pressure is the higher, the larger the number of molecules striking the wall over a certain time interval and the higher the velocities of impinging molecules.

The emergence of gas pressure can be explained with the help of a simple mechanical model. We take a disc (it plays the role of the manometer membrane) and fix it on a rod in the vertical position so that it can rotate together with the rod about a vertical axis (Fig. 19). Using an inclined spout, we direct a stream of small pellets onto the disc (the pellets play the role of molecules). As a result of multiple impacts of the pellets against the disc, a certain average force acts on it, which causes the turning of the rod and the bending of an elastic plate P. The effect of the impacts of individual pellets is, however, not noticeable.

THERMAL MOTION OF MOLECULES. At the very outset, we should give up the attempts to trace the motion of all the molecules constituting a gas. Their number is too large and their paths are very complicated. But there is no need to know how every molecule moves. We must find the result of the collective motion of all the gas molecules.

---

[1] It should be recalled that pressure is defined as the ratio of the magnitude F of the force acting normally to the surface, to the area S of this surface: $p = F/S$. Pressure is expressed in pascals or millimetres of mercury column: 1 Pa = 1 N/m^2 = 7.5×10^{-3} mm Hg.
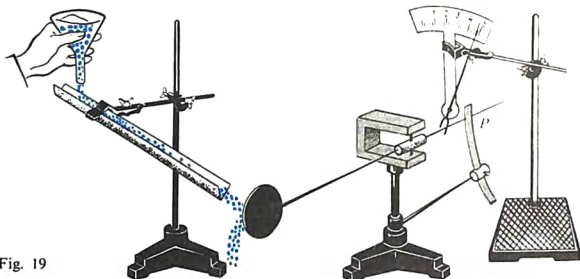
Fig. 19

The nature of motion of the whole aggregate of gas molecules is known from experiment (see Sec. 1.3). *Molecules are in a random (thermal) motion.* This means that the velocity of any molecule may turn out to be very high or very low, while the direction of their motion incessantly changes upon their mutual collisions.

Although the velocities of individual molecules may acquire various values, the AVERAGE VALUE of the magnitude of their velocity is quite definite. This is clear from the following simple example. The height of the pupils in a class is not the same, but there is a definite average value for this quantity. To find it, we must sum up the heights of the pupils and divide the result by their number.

THE MEAN SQUARE VELOCITY. In the further analysis, we shall have to know the square of the average velocity rather than the mean velocity. This quantity determines the average kinetic energy of molecules. It will be shown below that the average kinetic energy plays a very important role in the entire molecular kinetic theory.

We denote the magnitudes of the velocities of individual molecules by v_1, v_2, v_3, ..., v_N. The mean square velocity is determined by the following formula:

$$\overline{v^2} = \frac{v_1^2 + v_2^2 + v_3^2 + \dots + v_N^2}{N}, \tag{1.8}$$

where N is the number of molecules in a gas.

But the square of the magnitude of any vector is equal to the sum of the squares of its projections on the coordinate axes Ox, Oy and Oz[1]:

$$v^2 = v_x^2 + v_y^2 + v_z^2. \tag{1.9}$$

---

[1] Figure 20 and Pythagoras' theorem can be used to prove this statement. It is known from the course of mechanics that for a plane motion $v^2 = v_x^2 + v_y^2$. Formula (1.9) is a generalization of this expression for the motion in the three-dimensional space.
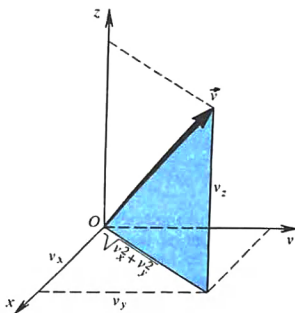
Fig. 20

The average values $\overline{v_x^2}$, $\overline{v_y^2}$ and $\overline{v_z^2}$ can be determined with the help of formulas similar to (1.8). The average value $\overline{v^2}$ and the average values of the squares of the projections are connected through a relation similar to (1.9):

$$\overline{v^2} = \overline{v_x^2} + \overline{v_y^2} + \overline{v_z^2}. \tag{1.10}$$

Indeed, relation (1.9) is valid for each molecule. Adding these relations for individual molecules and dividing both sides of the resultant relation by the number N of molecules, we arrive at formula (1.10).

Since the directions Ox, Oy and Oz are equivalent due to the random nature of molecular motion, the mean squares of the velocity projections are equal:

$$\overline{v_x^2} = \overline{v_y^2} = \overline{v_z^2}. \tag{1.11}$$

Using this relation, we substitute into (1.10) $\overline{v_x^2}$ for $\overline{v_y^2}$ and $\overline{v_z^2}$. Then the mean square of the projection of velocity is

$$\overline{v_x^2} = \frac{1}{3}\overline{v^2}, \tag{1.12}$$

i.e. the mean square of the velocity projection is equal to 1/3 of the mean square of the velocity itself. The factor of 1/3 appears since the space is three-dimensional, and hence any vector has three projections.

1.7. The Basic Equation of the Molecular Kinetic Theory of Gases

Let us suppose that a gas is contained in a vessel $ABCD$ of a rectangular cross section, such that one of its walls is a piston CD, that can move without friction (Fig. 21). The gas and the vessel are at the same temperature.

We calculate the gas pressure on the piston CD whose area is S. The surface of the piston is perpendicular to the axis Ox. The gas pressure
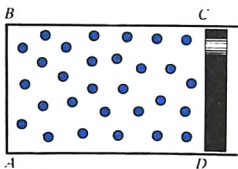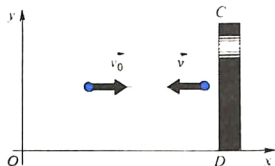
Fig. 21   Fig. 22

emerges as a result of impacts of molecules against the piston. To prevent the piston from being pushed out of the vessel, a certain force $\vec{F}$ must be applied to it from outside.

The derivation of the formula for the gas pressure is not very complicated but is rather cumbersome. We shall split it into four stages. Let us first calculate the impulse of the force exerted on the piston by a molecule during the impact. We shall then calculate the number of collisions of molecules with the piston over a time Δt. Multiplying these quantities and taking the average value of the square of the velocity of molecules, we shall determine the impulse of the average force exerted by all the molecules on the piston. Lastly, we divide the obtained value by the time Δt and by the area of the piston. As a result, we shall obtain an expression for the gas pressure.

COLLISION OF A MOLECULE WITH THE PISTON. Let us first consider the case when the velocity $\vec{v}_0$ of a molecule before the collision with the piston (Fig. 22) is perpendicular to the piston surface. In our model, the molecules are rigid balls. Upon collisions with a wall, they are bounced without a change in their kinetic energy. Such collisions are termed perfectly elastic. The magnitude of the velocity remains unchanged in this case, while the direction of motion is reversed: $\vec{v} = -\vec{v}_0$. The change in the momentum of the molecule is given by

$$m_0\vec{v} - m_0\vec{v}_0 = m_0\vec{v} - (-m_0\vec{v}) = 2m_0\vec{v}.$$

If the velocity of a molecule is at an arbitrary angle to the piston (Fig. 23), the projection v_{0x} of its velocity on the direction normal to the surface of the piston changes its sign ($v_x = -v_{0x}$), while the velocity projections v_{0y} and v_{0z} on the directions parallel to the surface of the piston remain unchanged: $v_y = v_{0y}$ and $v_z = v_{0z}$. A similar situation arises when a ball impinges at a smooth wall, if we assume that this collision is perfectly elastic.

The change in the projection of the momentum of a molecule on the x-axis is

$$m_0 v_x - m_0 v_{0x} = m_0 v_x - (-m_0 v_x) = 2m_0 v_x. \qquad (1.13)$$

According to the law of momentum conservation, the total momentum of the molecule and the piston remains unchanged. This means that the change in the momentum of the piston is equal in magnitude to the change in the momentum of the molecule. In other words, when a molecule strikes the

28
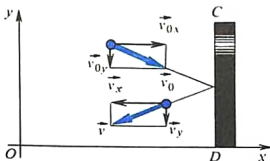
Fig. 23
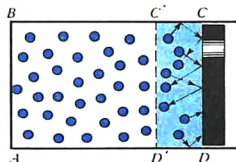
Fig. 24

piston, the magnitude of the momentum transferred to the piston is $2m_0 |v_x|$.

According to Newton's second law, the change in the momentum of a body is equal to the impulse of the force, viz. the product of the force by the time of its action. Therefore, the magnitude of the impulse of the force exerted on the piston by a molecule during the impact is equal to $2m_0 |v_x|$.

THE NUMBER OF COLLISIONS OF MOLECULES WITH THE PISTON. In order to calculate the impulse of the force exerted by all molecules on the piston, we have to calculate the number of collisions of molecules with the piston over a certain time interval Δt which is much longer than the time of collision of a single molecule with the piston.

During the time Δt, only those molecules can reach the piston that are at a distance not exceeding $CC' = |v_x|\Delta t$ from it (Fig. 24). The molecules at larger distances are unable to reach the piston. We must also take into account the fact that only the molecules with $v_x > 0$ can reach the wall CD during this time, i.e. the molecules moving from left to right.

The values of the velocity projections v_y and v_z do not affect the arrival of molecules at the piston CD. If a molecule undergoes an elastic collision with the wall BC or AD (Fig. 24), the velocity projection v_x remains unchanged, and the molecule still will be displaced along Ox by the segment $|v_x|\Delta t$.

The volume $CC'D'D$ is equal to $|v_x|\Delta t S$. If the number density of molecules, i.e. the number of molecules per unit volume is $n = N/V$, the number of molecules in the isolated volume is $n|v_x|\Delta t S$.

Due to the random nature of the molecular motion, only half the molecules in the isolated volume have the velocity projection $v_x > 0$ and move from left to right. For the other half of molecules $v_x < 0$, and they move in the opposite direction.

Consequently, the number z of collisions of molecules with the piston over the time Δt is equal to half the number of molecules in the isolated volume:

$$z = \frac{n}{2} |v_x| \Delta t S. \qquad (1.14)$$

THE IMPULSE OF THE AVERAGE FORCE EXERTED BY ALL MOLECULES ON THE PISTON. Every molecule colliding with the piston changes its momentum by $2m_0 |v_x|$. During the time Δt, all z molecules change its momentum by $2m_0 |v_x| z$. According to Newton's second law, the

29

magnitude $F\,\Delta t$ of the impulse of the force acting on the piston is equal to the magnitude of the change in its momentum:

$$F\,\Delta t = z\,2m_0\,|\,v_x\,| = nm_0 S v_x^2\,\Delta t. \tag{1.15}$$

We now must take into consideration the following circumstance. We argued as if all the molecules had the same velocity. Actually, the velocities of molecules are different, and each molecule makes its own contribution to pressure upon its collision with the piston. To take this into account, we must take the value of the square of the velocity projection, $\overline{v_x^2}$, averaged over all the molecules, instead of v_x^2.

The average value of the magnitude of the impulse of force is given by $\overline{F}\,\Delta t = m_0 n S \overline{v_x^2}\Delta t$. Since $\overline{v_x^2} = \overline{v^2}/3$ (see (1.12)), we have

$$\bar{F}\,\Delta t = \frac{1}{3} m_0 n S \overline{v^2}\,\Delta t. \tag{1.16}$$

GAS PRESSURE. Dividing both sides of Eq. (1.16) by the product $\Delta t S$, we can find the gas pressure:

$$p = \frac{1}{3} m_0 n \overline{v^2}. \tag{1.17}$$

This is the basic equation of the molecular kinetic theory.

Formula (1.17) connects a macroscopic quantity, viz. the pressure that can be measured by a manometer, with microscopic quantities characterizing molecules and is a sort of bridge between the two worlds: macroscopic and microscopic.

If we denote by $\bar{E}$ the mean kinetic energy of translatory motion of a molecule, $\bar{E} = m_0 \overline{v^2}/2$, then Eq. (1.17) can be written in the form

$$p = \frac{2}{3} n \bar{E}. \tag{1.18}$$

The pressure of an ideal gas is proportional to the product of the number of molecules per unit volume and the mean kinetic energy of the translatory motion of a molecule.

In the next chapter it will be shown that the mean kinetic energy of molecules is determined by the gas temperature.

?   1. What is an ideal gas in the molecular kinetic theory?
2. What is the mechanism of the emergence of gas pressure from the point of view of the molecular kinetic theory?
3. What is the mean value of the x-projection of the molecular velocity equal to?
4. Give the definition of the mean square velocity of molecules.
5. By what amount does the momentum of a molecule change upon its collision with the vessel wall?
6. What quantities determine the number of collisions of molecules with a piston of area S over a time Δt?
7. Write the basic equation of the molecular kinetic theory.

Problems with Solutions

To solve most of the problems concerning Chapter 1, you must be able to determine the molar masses of substances. For this purpose, the relative molecular mass must be determined from the relative atomic masses given in the Periodic Table of the Elements. The molar mass is then determined by the formula $M = 10^{-3} M_r$ kg/mole, where M is the molar mass and M_r is the relative molecular mass.

In some problems, the amount of substance or the number of molecules (atoms) in it has to be determined from the known mass of a body. For this purpose, the following formulas are used: $\nu = m/M$ and $N = (m/M)N_A$. The masses of individual molecules are determined by the formula $m_0 = M/N_A$. In some problems, the mass of a substance has to be expressed in terms of its density ρ and volume V.

These formulas, as well as all other formulas from Sec. 1.2 should be kept at hand while solving problems.

In a number of problems, the basic equation of the molecular kinetic theory in the form (1.17) or (1.18) should be used.

1. Determine the molar mass of water.

Solution. The relative atomic masses of hydrogen and oxygen are 1.00797 and 15.9994 respectively. The chemical formula of water is H_2O. Consequently, the relative molecular mass of water is

$$M_r = 2 \times 1.00797 + 15.9994 = 18.01534 \simeq 18.$$

The molar mass of water is $M \simeq 10^{-3} \times 18$ kg/mole $\simeq 0.018$ kg/mole.

2. Determine the amount of substance and the number of molecules contained in 1 kg of carbon dioxide.

Solution. Since the molar mass of carbon dioxide is $M = 0.044$ kg/mole, the amount of substance $\nu = m/M = 1/0.044$ mole $\simeq 22.7$ mole. The number of molecules is $N = (m/M)N_A = \nu N_A = 22.7 \times 6.02 \times 10^{23} \simeq 1.4 \times 10^{25}$.

3. A vessel having a volume $V = 0.01$ m^3 contains $N = 2.7 \times 10^{27}$ gas molecules. The pressure on the vessel walls is $p = 10^5$ Pa. What is the mean kinetic energy $\bar{E}$ of the translatory motion of the molecules?

Solution. According to the basic equation of the molecular kinetic energy of gases in the form (1.18), $p = (2/3)n\bar{E}$. The number density of molecules is $n = N/V = 2.7 \times 10^{25}$ m^{-3}. Hence

$$\bar{E} = \frac{3}{2}\frac{p}{n} = \frac{3 \times 10^5 \text{ N/m}^2}{2 \times 2.7 \times 10^{25} \text{ m}^{-3}} \simeq 5.6 \times 10^{-21} \text{ J}.$$

4. The density of a gas in a gas-filled electric lamp is $\rho = 0.9$ kg/m^3. After the lamp has been switched on, the pressure in it increases from $p_1 = 8 \times 10^4$ Pa to $p_2 = 1.1 \times 10^5$ Pa. What is the increase in the mean velocity of gas molecules in this case?

Solution. The product of the mass m_0 of an individual molecule by the molecular number density (the number of molecules per unit volume) n is

31

equal to the mass of the molecules contained in a unit volume, i.e. to the gas density $\rho = m_0 n$. Consequently, the basic equation (1.17) of the molecular kinetic theory can be written in the form of $p = (1/3)\rho v^2$.

Therefore,

$$\bar{v}_1 = \sqrt{\frac{3p_1}{\rho}}, \quad \bar{v}_2 = \sqrt{\frac{3p_2}{\rho}}.$$

Hence

$$\bar{v}_2 - \bar{v}_1 = \sqrt{\frac{3}{\rho}}(\sqrt{p_2} - \sqrt{p_1}) \simeq 90 \text{ m/s}.$$

Exercise 1
1. What area can be occupied by a drop of olive oil of volume 0.02 cm^3 when it spreads over a surface of water?
2. Determine the molar masses of hydrogen and helium.
3. What is the ratio of the number of atoms in 12 kg of carbon and the number of molecules in 16 kg of oxygen?
4. What amount of substance (in moles) is contained in 1 g of water?
5. What is the number of molecules in 10 g of oxygen?
6. The molar mass of nitrogen is 0.028 kg/mole. What is the mass of a nitrogen molecule?
7. Determine the number of atoms in 1 m^3 of copper if its molar mass $M = 0.0635$ kg/mole and its density $\rho = 9000$ kg/m^3.
8. The density of diamond is 3500 kg/m^3. What volume is occupied by 10^{22} atoms of this substance?
9. What is the change in the gas pressure if the number density of molecules in the gas has increased thrice, while the mean molecular velocity has become equal to 1/3 of its initial value?
10. What is the pressure of a gas in a vessel if the mean square velocity of its molecules is $\overline{v^2} = 10^6$ m^2/s^2, the molecular number density $n = 3 \times 10^{25}$ m^{-3}, and the mass of each molecule is $m_0 = 5 \times 10^{-26}$ kg?
11. A flask whose volume is 1.2 l contains 3×10^{22} helium molecules. What is the mean kinetic energy of each molecule? The pressure in the flask is 10^5 Pa.
12. Calculate the mean square velocity of molecules in a gas if its mass $m = 6$ kg, volume $V = 4.9$ m^3 and pressure $p = 200$ kPa.

Main Points of Chapter 1

According to the basic concepts of the molecular kinetic theory, all bodies consist of molecules (or atoms). Repulsive forces act among molecules at distances smaller than the molecular dimensions, while attractive forces operate at large distances. Molecules take part in random thermal motion.

The masses of molecules are small, and the number of molecules in large (macroscopic) bodies is very large. For this reason, the masses of molecules and their number in macroscopic bodies are expressed in relative units.

The relative molecular (or atomic) mass is the ratio of the mass of

a molecule (or atom) to 1/12 of the mass of a carbon atom (1.2).

The amount of substance is equal to the ratio of the number of molecules in a given body to Avogadro's number, i.e. the number of molecules in a mole of a substance.

The amount of substance is expressed in moles. A mole is the amount of substance containing the same number of molecules as in 0.012 kg of carbon.

The number of molecules in a mole of a substance is known as Avogadro's number $N_A = 6.02 \times 10^{23}$ mole^{-1}.

The molar mass is the mass of one mole of a substance (1.4).

The separation between molecules in gases is much larger then their dimensions. The molecules (or atoms) in liquids and solids are in the immediate proximity to one another. Atoms (or molecules) in solids are arranged in a strict order and vibrate about invariable equilibrium positions. In liquids, molecules are arranged chaotically and from time to time jump from one equilibrium position to another.

An ideal gas in the molecular kinetic theory is a gas consisting of molecules with a negligibly weak interaction.

According to the basic equation of the molecular kinetic theory, the gas pressure is proportional to the product of the molecular number density and the mean kinetic energy of the translatory motion of a molecule:

$$p = \frac{2}{3} n\bar{E}, \quad \text{where} \quad \bar{E} = \frac{m_0 \overline{v^2}}{2}.$$

2 TEMPERATURE. ENERGY OF THERMAL MOTION OF MOLECULES

2.1. Thermal Equilibrium. Temperature

MACROSCOPIC PARAMETERS. There is no need to resort each time to the molecular kinetic theory for describing processes occurring in gases and other macroscopic bodies. The behaviour of macroscopic bodies, and in particular gases, can be characterized by a small number of physical quantities pertaining to molecules in aggregate rather than to individual molecules constituting bodies. These quantities include the volume V, pressure p and temperature t.

Suppose, for example, a given mass of a gas occupies a certain volume and has certain temperature and pressure. Volume and pressure are mechanical quantities describing the state of the gas. Temperature is not considered in mechanics since it characterizes the internal state of the body.

Any macroscopic body or a group of macroscopic bodies is called a THERMODYNAMIC SYSTEM. The quantities (V, p, t) characterizing the state of the thermodynamic system without taking into account the molecular structure of bodies are known as MACROSCOPIC (or THERMODYNAMIC) PARAMETERS. Macroscopic parameters are not confined just to volume, pressure and temperature. For example, for a gas mixture we must also know the concentrations of individual components of the mixture. Ordinary atmospheric air is an example of such a mixture.

COLD AND HOT BODIES. The concept of temperature occupies the central place in the theory of thermal phenomena. We all know very well the difference between cold and hot bodies. We can determine by touch which of the bodies is hotter, and we say that this body has a higher temperature. The temperature characterizes the DEGREE OF HOTNESS of a body (cold, warm, or hot). The temperature is measured by a special instrument called the thermometer. Its construction is based on the property of bodies to change their volume upon heating or cooling.

THERMAL EQUILIBRIUM. To measure the temperature of a human body, a clinical thermometer should be held in the armpit for 5-8 min. During this time, mercury in the thermometer is heated and its level rises. The length of the mercury column indicates the temperature of the body. In

the same way, the temperature of any other body can be measured by any type of thermometer. A thermometer never indicates the temperature of a body immediately after it was brought in contact with the body. A certain time is required before the temperatures of the body and the thermometer level out and THERMAL EQUILIBRIUM is established between them, after which the temperature does not change.

Thermal equilibrium is established with time between any bodies at different temperatures. Drop a piece of ice into a glass with water and tightly close the glass by a lid. The ice will melt and the water will cool. After the ice has melted, water starts getting heated. After the water has acquired the temperature of the surrounding air, the temperature of the glass with water does not change any more.

This and other similar simple observations lead to the conclusion about a very important general property of thermal phenomena. Any macroscopic body or a group of bodies (a thermodynamic system) *under invariable external conditions spontaneously reaches the state of thermal equilibrium.* Thermal, or thermodynamic, equilibrium is a state in which all the macroscopic parameters remain unchanged for an infinitely long time. This means that the volume and pressure of the system do not change, there is no heat exchange, no mutual conversion of gases, liquids and solids is observed, and so on. In other words, the temperature of the system remains unchanged.

However, microscopic processes within a body do not cease even in a thermal equilibrium. As a result of collisions the positions of molecules keep on changing as well as their velocities.

TEMPERATURE. A thermodynamic system can be in different states of thermal equilibrium. In each of these states, the temperature has its own, strictly definite value. Other quantities may have different (but constant) values in a state of thermal equilibrium. For example, the volumes of different parts of a system, as well as their pressures, may be different if there are rigid partitions between them. If you bring a ball filled with compressed air from the street into a room, the temperature of the air in the ball and in the room equalizes in a certain time. The air pressure in the ball still remains higher than the pressure in the room.

The temperature characterizes the state of thermal equilibrium of a macroscopic system: the temperature has the same value in all parts of the system in thermal equilibrium.

If two bodies are at the same temperature, there is no heat exchange between them. The bodies are in thermal equilibrium in this case. If, however, the temperatures of the bodies are different, these bodies exchange energy when brought in thermal contact. The body at a higher temperature gives energy to the body at a lower temperature. *The temperature difference indicates the direction of heat exchange between the bodies.*

MEASUREMENT OF TEMPERATURE. THERMOMETERS. The measurement of temperature can be based on the temperature variation of any macroscopic quantity: volume, pressure, electric resistance and so on.

In practice, the dependence of the volume of a liquid (mercury or alcohol) on temperature is used most frequently. While graduating a thermometer, the

Fig. 25

temperature of melting ice is taken as a reference point (0), while the other reference point (100) is the temperature of boiling water under the normal atmospheric pressure (Celsius scale). Between the points 0 and 100, the scale is divided into 100 equal parts called degrees (Fig. 25). The displacement of the liquid column by one division corresponds to a change in temperature by 1 °C.

Since different liquids expand differently upon heating, the scale obtained in this way will be determined to a certain extent by the properties of the liquid. Naturally, 0 and 100 °C will be the same for all thermometers, but, say, the divisions corresponding to 50 °C will not coincide.

What substance should then be chosen to get rid of this dependence? It was noted that unlike liquids, all rarefied gases (hydrogen, helium and oxygen) expand in the same way upon heating and change their pressure with temperature identically. For this reason, the change in pressure of a certain amount of a rarefied gas at a constant volume or the change in the gas volume at a constant pressure is used in physics for establishing a rational temperature scale. This scale is sometimes called the ideal gas temperature scale. Using this scale, we can get rid of another considerable drawback of the Celsius scale, viz. the arbitrariness in the choice of the reference point, or the zero temperature. The reference point could be taken on equal rights at the temperature of boiling water.

We shall consider in detail how rarefied gases can be used for determining temperature.

? 1. What quantities characterize the state of macroscopic bodies?
2. What are the salient features of the state of thermal equilibrium?
3. Give examples of establishing thermal equilibrium of bodies surrounding us in everyday life.
4. Which physical quantity characterizes the state of thermal equilibrium?
5. What properties of macroscopic bodies are used for measuring temperature?
6. What are the advantages of using rarefied gases for the measurement of temperature?

2.2. Temperature of a Gas

MEAN KINETIC ENERGY OF GAS MOLECULES AT THERMAL EQUILIBRIUM. In thermal equilibrium, all gases have the same temperature that does not depend on the kind of a gas. To determine the temperature we must find out which physical quantity has the property of temperature in the molecular kinetic theory.
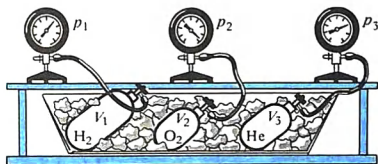
Fig. 26

It is known from the course of "Junior Physics" that the molecular velocity is the higher, the higher the temperature of a body. With heating the pressure in a closed vessel becomes higher. On the other hand, according to the basic equation (1.18) of the molecular kinetic theory, the pressure p is proportional to the mean kinetic energy of translatory motion of a molecule: $p = (2/3)n\bar{E}$. In thermal equilibrium, when the pressure of a given mass of a gas and its volume are fixed, the mean kinetic energy of the gas molecules must have a strictly definite value like the temperature.

It is natural to assume that in thermal equilibrium the mean kinetic energies of molecules of all gases are identical. Of course, it is just an assumption which should be verified experimentally. Such a verification cannot be carried out directly since it is very difficult to measure the mean kinetic energy of molecules. Using the basic equation in the molecular kinetic theory, we can, however, express it in terms of macroscopic parameters.

Since the number density of gas molecules is $n = N/V$, it follows from Eq. (1.18) that $p = (2/3)(N/V)\bar{E}$, or

$$\frac{pV}{N} = \frac{3}{2}\bar{E}. \tag{2.1}$$

Pressure and volume can be measured directly. The number of molecules can be determined from the mass m of the gas, Avogadro's number N_A and the molar mass M. According to formula (1.7), we have

$$N = \frac{m}{M} N_A.$$

If the kinetic energy $\bar{E}$ is indeed the same for all the gases in thermal equilibrium, the quantity pV/N should be the same for all gases. This assumption can be confirmed or refuted only by experiment.

GASES IN THERMAL EQUILIBRIUM. The experiment can be carried out as follows. Let us take several vessels filled with different gases, say, hydrogen, helium and oxygen. The vessels have definite volumes and have manometers attached to them. This allows us to measure pressure in each vessel. The masses of the gases, and hence the numbers of molecules in each vessel, are known.

Let us bring the gases into thermal equilibrium. For this purpose, we place them into melting ice and wait until the thermal equilibrium sets in and the gas pressure stops changing (Fig. 26). After that we can state that all the gases have the same temperature $0\,°C$. The pressures p, volumes V and the

numbers of molecules N are different for different gases. Let us calculate the quantity pV/N for hydrogen. If, for example, 1 mole of hydrogen occupies the volume $V_{H_2} = 0.1$ m^3, its pressure at 0 °C turns out to be $p_{H_2} = 22.65 \times 10^3$ Pa. Hence

$$\frac{p_{H_2} V_{H_2}}{N_{H_2}} = \frac{22.65 \times 10^3 \times 0.1}{6.02 \times 10^{23}} \frac{N}{m^2} \ m^3 = 3.76 \times 10^{-21} \ J. \tag{2.2}$$

The ratio of the product of the gas pressure and its volume to the number of molecules in it has the same value for other gases also at the temperature of melting ice. We denote this ratio by Θ_0. Then

$$\frac{p_{H_2} V_{H_2}}{N_{H_2}} = \frac{p_{He} V_{He}}{N_{He}} = \frac{p_{O_2} V_{O_2}}{N_{O_2}} = \Theta_0. \tag{2.3}$$

Thus, our assumption turns out to be true.

As a matter of fact, relation (2.3) is not absolutely accurate. Under pressures of thousands of atmospheres when gases become very dense, the ratio pV/N is no longer strictly definite and independent of the volume occupied by a gas. It is satisfied for sufficiently rarefied gases which can be treated as ideal gases.

If we immerse all the vessels with the gases into boiling water (at 100 °C), under the normal atmospheric pressure, $pV/N = \Theta_{100}$ remains as before the same for all the gases, but its magnitude becomes larger. Experiment shows that now we have

$$\frac{pV}{N} = \Theta_{100} = 5.14 \times 10^{-21} \ J. \tag{2.4}$$

DETERMINATION OF TEMPERATURE. Thus we can state that the value of Θ increases with temperature. Moreover, Θ depends only on temperature. Indeed, for rarefied gases Θ is independent of the kind of a gas, its volume and pressure, the number of particles in a vessel, and the shape of the vessel itself. This experimental result allows us to treat Θ as a natural measure of temperature determined through other macroscopic parameters of a gas. It could be possible, in principle, to consider the very quantity Θ as temperature and to measure temperature in the energy units, viz. joules. However, these units are inconvenient for practical application, since a very small quantity of the order of 10^{-21} J would correspond to 100 °C. Secondly, and this is the main reason, the temperature has been conventionally measured in degrees for a long time.

2.3. Absolute Temperature. Temperature as a Measure of the Mean Kinetic Energy of Molecules

Instead of the temperature Θ which is measured in the energy units, we introduce a temperature that is measured in degrees.

We shall assume that, by definition, the value of Θ is directly proportional

Ludwig Eduard Boltzmann (1844–1906) was a great Austrian physicist, one of the founders of the molecular kinetic theory. In his works, the molecular kinetic theory appeared for the first time as a logically elegant and consistent physical theory. Boltzmann contributed much to the development and popularization of the Maxwell theory of electromagnetic field. Fighter by nature, he passionately defended the necessity of the molecular interpretation of thermal phenomena and bore the brunt in the struggle with scientists who denied the existence of molecules.

to temperature T measured in degrees[1]:

$$\Theta = kT, \tag{2.5}$$

where k is the proportionality factor. The temperature defined by Eq. (2.5) is called the ABSOLUTE TEMPERATURE. It will be shown later that this term is sufficiently justified.

Taking into account definition (2.5), we obtain

$$\frac{pV}{N} = kT. \tag{2.6}$$

On the basis of this formula, a temperature scale independent of the substance used for measuring temperature is established.

ABSOLUTE TEMPERATURE ZERO. Obviously, the temperature defined by formula (2.6) cannot be negative since all the quantities appearing on the left-hand side of (2.6) are definitely positive.

Consequently, the smallest possible value of temperature T is $T = 0$, when either pressure p or volume V is equal to zero. The limiting temperature at which the pressure of an ideal gas vanishes at a fixed volume or the volume of an ideal gas tends to zero at a constant pressure is called the ABSOLUTE ZERO OF TEMPERATURE. This is the lowest temperature in nature, that "extreme or ultimate degree of cold" whose existence was predicted by Lomonosov.

ABSOLUTE TEMPERATURE SCALE. The English scientist W. Kelvin introduced the absolute temperature scale. THE ZERO TEMPERATURE ON THE ABSOLUTE SCALE (also known as the thermodynamic temperature scale or

---

[1] We actually assume with such a definition of temperature that the pressure at constant volume and the volume at constant pressure increase linearly with temperature measured in degrees centigrade (C), i.e. increase in the same way as when the temperature is measured in the energy units.

the Kelvin scale) CORRESPONDS TO THE ABSOLUTE ZERO, AND EACH UNIT OF TEMPERATURE ON THIS SCALE IS EQUAL TO A DEGREE ON THE CELSIUS SCALE.

The SI unit of absolute temperature is a kelvin (denoted by K).

BOLTZMANN'S CONSTANT. Let us define the coefficient k in formula (2.6) so that a kelvin is equal to a degree on the Celsius scale.

We know the values of Θ at $0\,°C$ and $100\,°C$ (formulas (2.2) and (2.4)). Let us denote the absolute temperature corresponding to $0\,°C$ and to $100\,°C$ by T_1 and T_2 respectively. Then

$$\Theta_{100} - \Theta_0 = k(T_2 - T_1)\,K,$$

or

$$\Theta_{100} - \Theta_0 = k \times 100\ K = (5.14 - 3.76) \times 10^{-21}\ J.$$

Hence

$$k = \frac{5.14 - 3.76}{100} \times 10^{-21}\ J/K = 1.38 \times 10^{-23}\ J/K.$$

The coefficient

$$\underline{k = 1.38 \times 10^{-23}\ J/K} \tag{2.7}$$

is known as BOLTZMANN'S CONSTANT after the great Austrian physicist L. Boltzmann, one of the founders of the molecular kinetic theory of gases.

Boltzmann's constant connects the temperature Θ measured in energy units with the temperature T measured in kelvins. This is one of the most important constants in the molecular kinetic theory.

RELATION BETWEEN THE ABSOLUTE SCALE AND THE CELSIUS SCALE. If we know Boltzmann's constant, we can determine the value of the absolute zero on the Celsius scale. For this we shall first find the value of the absolute temperature corresponding to $0\,°C$.

Since $kT_1 = 3.76 \times 10^{-21}\ J$ for $0\,°C$, we get

$$T_1 = \frac{3.76 \times 10^{-21}}{1.38 \times 10^{-23}}\ K \simeq 273\ K.$$

One kelvin and one degree centigrade coincide. Therefore, any value of the absolute temperature T is 273 degrees higher than the corresponding temperature t on the Celcius scale:

$$T = t + 273. \tag{2.8}$$

But the change ΔT in the absolute temperature is equal to the change Δt on the Celcius scale: $\Delta T = \Delta t$.

In Fig. 27, the absolute scale and the Celsius scale are compared. The absolute zero corresponds to the temperature $t = -273\,°C$.[1]

TEMPERATURE AS A MEASURE OF THE MEAN KINETIC ENERGY OF MOLECULES. The following important conclusion follows

---

[1] A more accurate value of the absolute zero is $-273.15\,°C$.
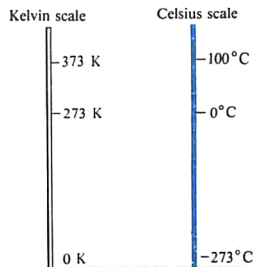
Fig. 27

from the basic equation in the molecular kinetic theory in the form (2.1) and the definition (2.6) of temperature: the absolute temperature is a measure of the mean kinetic energy of molecular motion. Let us prove this statement.

The left-hand sides of the equations $pV/N = (2/3)\bar{E}$ and $pV/N = kT$ are identical. Consequently, their right-hand sides must be equal. This leads to the relation between the mean kinetic energy of the translatory motion of a molecule and the temperature:

$$\bar{E} = \frac{3}{2} kT. \tag{2.9}$$

The mean kinetic energy of chaotic translatory motion of gas molecules is proportional to absolute temperature. The higher the temperature, the higher the velocity of molecules. Thus, the above hypothesis about the connection between the temperature and the average velocity of molecules has found a reliable substantiation.

Relation (2.9) between the temperature and the mean kinetic energy of translatory motion of molecules has been established for ideal gases. However, it turns out to be valid for any substance in which the motion of atoms or molecules is governed by the laws of Newtonian mechanics. It is satisfied for liquids and solids whose atoms can only vibrate about equilibrium positions at the sites of the crystal lattice.

As the temperature approaches the absolute zero, the energy of thermal motion of molecules also tends to zero.

DEPENDENCE OF THE GAS PRESSURE ON TEMPERATURE AND THE NUMBER DENSITY OF ITS MOLECULES. Considering that $N/V = n$, we obtain from (2.6) the following expression connecting the gas pressure with the number density of gas molecules and the temperature:

$$p = nkT. \tag{2.10}$$

It follows from this formula that at the same temperature and pressure, the number density of molecules is the same for all gases.

This formula also leads to Avogadro's law known from the course of

chemistry: at the same pressure and temperature equal volumes of all gases contain the same number of molecules.

1. What is the relation connecting the volume, pressure and the number of molecules of different gases in thermal equilibrium?
2. What is the physical meaning of Boltzmann's constant?
3. What are the advantages of the absolute temperature scale over other scales?
4. Which temperature on the Celsius scale corresponds to the absolute zero?
5. What is the relation between the temperature and the mean kinetic energy of translatory motion of molecules?
6. What is the relation connecting the pressure of a gas with its temperature and the number density of its molecules?
7. What is the physical meaning of the absolute zero temperature?

2.4. Measurement of Velocities of Gas Molecules

AVERAGE VELOCITY OF THERMAL MOLECULAR MOTION. Equation (2.9) makes it possible to calculate the mean square velocity of molecular motion. Substituting into this equation $\bar{E} = m_0 \overline{v^2}/2$, we obtain the expression for the mean square velocity:

$$\overline{v^2} = 3kT/m_0. \tag{2.11}$$

The square root of this quantity is known as the root-mean-square velocity:

$$\bar{v} = \sqrt{3kT/m_0}. \tag{2.12}$$

Using this formula to calculate, for example, the velocity of nitrogen molecules at $t = 0\,°C$, we obtain $\bar{v} \simeq 500$ m/s. Hydrogen molecules have a velocity $\bar{v} \simeq 1800$ m/s at the same temperature.

When these values were obtained for the first time (in the second half of the 19th century), many physicists were stunned. According to calculations, the velocities of gas molecules turned out to be higher than the velocities of gun shells! This circumstance even gave grounds for doubting the validity of the molecular kinetic theory. It is well known, for example, that odours propagate at a slow rate: it takes about tens of seconds for the scent of perfume poured out in one corner of the room to reach the other corner. This fact can be easily explained. As a result of collisions, the path of each molecule is a complicated broken line (Fig. 28). The molecule has a large velocity on rectilinear segments of its polygonal path. However, the advance of the molecule in a certain direction is on the average small even over a time of the order of several minutes. As the molecule goes over from point A to point B, the path traversed by it is much longer than the distance AB.

EXPERIMENTAL DETERMINATION OF MOLECULAR VELOCITIES. The validity of formula (2.12) was proved in the experiments on the determination of velocities of molecules. One such experiment was carried out by Stern in 1920.
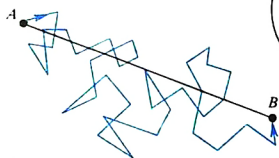
Fig. 28
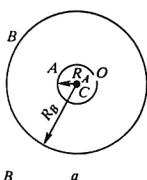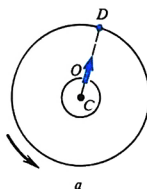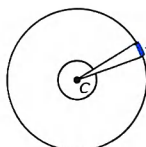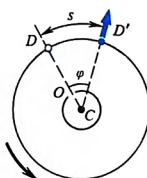
Fig. 29

Fig. 30

The experimental set-up consists of two coaxial cylinders A and B rigidly connected to each other (Fig. 29a). The cylinders can be rotated at a constant frequency. A thin platinum wire C coated by a silver layer is stretched along the axis of the smaller cylinder. An electric current is passed through the wire and a narrow slit O is cut in the wall of the cylinder A. The air is pumped out of the cylinders. The cylinder B is kept at room temperature.

At the beginning, the cylinders are at rest. As the current passes through the wire, the silver vaporizes and the inner cylinder is filled by the gas of silver atoms. Some of them fly through the slit O and are deposited on the inner surface of the cylinder B. As a result, a narrow silver strip D is formed just against the slit (Fig. 29b).

Then the cylinders start to rotate at a speed of n rps. During the time t required for a silver atom to cover the distance $R_B - R_A$ (the difference between the radii of the cylinders), the cylinders are turned by an angle φ. As a result, the atoms moving at a constant velocity get to the inner surface of the larger cylinder not against the slit O (Fig. 30a) but at a certain distance s from the end of the radius passing through the middle of the slit (Fig. 30b), since the atoms move in a straight line.

If we denote by v_B the magnitude of the linear velocity of the point on the surface of the external rotating cylinder, we get

$$s = v_B t = 2\pi n R_B t. \tag{2.13}$$

As a matter of fact, not all silver atoms have the same velocity. Therefore the distances s are slightly different for different atoms. The distance s should be treated as the distance between the regions corresponding to the highest

43

concentration of silver atoms on strips D and D'. This distance corresponds to the average velocity of silver atoms.

The average velocity of a silver atom is

$$\bar{v} = \frac{R_B - R_A}{t}.$$

Substituting into this formula the value of t from (2.13), we obtain

$$\bar{v} = \frac{2\pi n R_B (R_B - R_A)}{s}.$$

Knowing n, R_A and R_B·and measuring the average displacement s of the silver strip due to the rotation of the cylinders, we can find the average velocity of silver atoms.

The experimental values of velocities coincide with the theoretical value of the root-mean-square velocity. *This serves as an experimental verification of the validity of formula* (2.12), *and hence* (2.9), according to which the mean kinetic energy of a molecule is proportional to the absolute temperature.

AVERAGE VELOCITY OF A BROWNIAN PARTICLE. Formula (2.12) explains why the intensity of Brownian movement increases with temperature of a liquid and decreases with increasing mass of a particle. The Brownian particle takes part in the thermal motion of molecules. Therefore, its mean kinetic energy is also determined by formula (2.9), while its root-mean-square velocity is given by

$$\bar{v} = \sqrt{3kT/m},$$

where m is the mass of the Brownian particle. If the mass of the particle is large, the average velocity of its motion is so small that the motion of the particle cannot be observed.

?

1. What is the change in the root-mean-square velocities of molecules due to a fourfold increase in the temperature?
2. Which molecules have higher velocities in the atmosphere: nitrogen molecules or oxygen molecules?
3. Why is the thickness of the silver layer on the inner surface of the rotating cylinder in the Stern experiment not uniform over the width of the strip?

Problems with Solutions

Problems in this chapter are solved with the help of formula (2.6) (which defines the absolute temperature), formula (2.9) (which relates the mean energy of random motion with temperature) and formula (2.12) for the root-mean-square velocity of molecules. Some problems can be conveniently solved by using formula (2.10) relating the gas pressure with the number density of molecules and the absolute temperature.

The value of Boltzmann's constant (2.7) will also be needed.

1. Find the value of the ratio of the product of the gas pressure and its volume to the number of molecules at $t = 300°C$?

Solution. According to formula (2.6), $pV/N = kT$, where $k = 1.38 \times 10^{-23}$ J/K is Boltzmann's constant. Since the absolute temperature $T = t + 273 = 573$ K, we obtain

$$\frac{pV}{N} = 1.38 \times 10^{-23} \text{ J/K} \times 573 \text{ K} = 7.9 \times 10^{-21} \text{ J}.$$

2. Determine the root-mean-square velocity of a gas molecule at $0\,°C$, if the molar mass of the gas is $M = 0.019$ kg mole.

Solution. The root-mean-square velocity of molecules is calculated with the help of formula (2.12). Considering that $m_0 = M/N_A$ and $T = 273$ K, we obtain

$$\bar{v} = \sqrt{3kT/m_0} = \sqrt{3kN_A T/M}$$
$$= \sqrt{\frac{3 \times 1.38 \times 10^{-23} \times 6.02 \times 10^{23} \times 273}{0.019}} \text{ m/s} \simeq 600 \text{ m/s}.$$

3. A certain amount of hydrogen is at a temperature $T_1 = 200$ K and a pressure $p_1 = 400$ Pa. The gas is heated to $T_2 = 10\,000$ K at which the molecules of hydrogen are decomposed into atoms almost completely. Determine the new value of the gas pressure p_2 if its volume and mass remain unchanged.

Solution. According to formula (2.10), at the temperature T_1 the gas pressure is $p_1 = n_1 k T_1$, where n_1 is the number density of hydrogen molecules.

When a hydrogen molecule splits into atoms, the number density of particles in the vessel increases twice. Consequently, the number density of hydrogen atoms is $n_2 = 2n_1$. The pressure of atomic hydrogen is $p_2 = n_2 k T_2 = 2n_1 k T_2$.

Dividing the second equation by the first termwise, we obtain $p_2/p_1 = 2T_2/T_1$, whence $p_2 = 2p_1 T_2/T_1 = 40$ kPa.

Exercise 2
1. What would be the value of Boltzmann's constant if the SI unit of temperature, kelvin, were equal to $2\,°C$ instead of $1\,°C$?
2. What is the mean kinetic energy of an argon molecule at a temperature of $17\,°C$?
3. Modern vacuum pumps make it possible to decrease pressure to 1.3×10^{-10} Pa (10^{-12} mm Hg). What number of gas molecules is contained at this pressure and at $27\,°C$ in 1 cm^3?
4. Compare the numbers of molecules contained in a room having a volume of 50 m^3 under normal atmospheric pressure at $20\,°C$ and in a glass of water whose volume is 200 cm^3.
5. The room-mean-square velocity of a molecule of a gas is 540 m/s at a temperature of $100\,°C$. Find the mass of the molecule.
6. Determine the increase (in percent) in the root-mean-square velocity of a water molecule in the blood of a patient upon an increase in the temperature of his body from 37 to $40\,°C$.

Main Points of Chapter 2

The internal state of macroscopic bodies is determined by quantities known as the macroscopic parameters. They include pressure, volume and temperature. The temperature is a measure of the intensity of the thermal motion of molecules, which characterizes the state of thermal equilibrium of a thermodynamic system. In thermal equilibrium, no macroscopic processes occur and all the macroscopic parameters remain unchanged, while the temperature has the same value in all parts of the system.

Temperature is measured by thermometers. In the construction of any thermometer, the variation of any thermodynamic parameter with temperature is used.

In all rarefied gases, unlike liquids and solids, the volume varies in the same way with temperature at a constant pressure (or pressure varies at a constant volume) upon heating. For this reason, rarefied gases are used for establishing a temperature scale (gas temperature scale).

Experiments show that the quantity pV/N is the same for all gases in thermal equilibrium and depends only on temperature. Therefore, it can be used for determining the absolute temperature T through the formula $pV/N = kT$, where $k = 1.38 \times 10^{-23}$ J/K is Boltzmann's constant. The absolute temperature is given by $T = t + 273$, where t is the temperature on the Celsius scale. The minimum temperature on the Kelvin scale is the absolute zero of temperature, equal to $-273\,°C$. On the absolute scale, the temperature is measured in kelvins. A degree kelvin is equal to a degree on the Celsius scale.

Comparing the expression for pressure given by the basic equation in the molecular kinetic theory with the formula $pV/N = kT$ defining the temperature, we obtain a very important result: the mean kinetic energy of random motion of molecules is proportional to the absolute temperature: $\bar{E} = (3/2)kT$.

The root-mean-square velocity of molecules is calculated by the formula $\bar{v} = \sqrt{3kT/m_0}$ and amounts to several hundred meters per second at 0 C.

3

EQUATION OF STATE FOR AN IDEAL GAS. GAS LAWS

3.1. Equation of State for an Ideal Gas

EQUATION OF STATE. We have considered in detail the behaviour of an ideal gas from the point of view of the molecular kinetic theory and established the dependence of the gas pressure on the number density of its molecules and the temperature (Formula (2.10)). This dependence can be used to obtain an equation connecting all the three macroscopic parameters (p, V and T) characterizing the state of a given mass of a sufficiently rarefied gas. This equation is known as the EQUATION OF STATE FOR AN IDEAL GAS.

Let us substitute the expression for the number density of gas molecules into Eq. (2.10) $p = nkT$. Taking into account formulas (1.3) and (1.7), the number density of gas molecules can be represented as follows:

$$n = \frac{N}{V} = \frac{1}{V} \frac{m}{M} N_A,$$

(3.1)

where N_A is Avogadro's number, m is the mass of the gas and M is its molar mass.

Substituting (3.1) into (2.10), we obtain

$$pV = \frac{m}{M} k N_A T.$$

(3.2)

The product of Boltzmann's constant k and Avogadro's number N_A is called the UNIVERSAL (MOLAR) GAS CONSTANT and is denoted by R:

$$R = kN_A = 1.38 \times 10^{-23} \times 6.02 \times 10^{23} \ \text{J/(mole} \cdot \text{K)} = 8.31 \ \text{J/(mole} \cdot \text{K)}.$$

(3.3)

Substituting into Eq. (3.2) the universal gas constant R for the product kN_A, we obtain the equation of state for an arbitrary mass of an ideal gas:

$$pV = \frac{m}{M} R T.$$

(3.4)
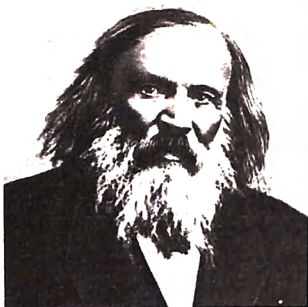
Dmitrii Ivanovich Mendeleev (1834–1907) was a great Russian scientist who created the Periodic System of Elements, one of the most deep-rooted generalizations in science. Mendeleev was the author of very important works on the theory of gases and mutual conversions of gases and liquids (he discovered the critical temperature above which a gas cannot be converted into a liquid). As a leading social worker, Mendeleev spared no efforts for the development of productive power of Russia, for utilization of mineral resources and for the advancement of chemical industry.

The only quantity in this equation which depends on the type of a gas is its molar mass.

The equation of state (3.4) leads to a relation connecting the pressure, volume and temperature of an ideal gas in any two states.

If we mark by subscript "1" the parameters characterizing the first state of a gas and by "2" the parameters corresponding to the second state, then according to Eq. (3.4) we obtain for a given mass of the gas

$$\frac{p_1 V_1}{T_1} = \frac{m}{M} R \text{ and } \frac{p_2 V_2}{T_2} = \frac{m}{M} R.$$

The right-hand sides of these equations are identical. Consequently, their left-hand sides must be equal:

$$\frac{p_1 V_1}{T_1} = \frac{p_2 V_2}{T_2} = \text{const.} \tag{3.5}$$

The equation of state in the form (3.5) is known as Clapeyron's equation [1] and is one of possible forms of the equation of state.

The equation of state in the form (3.4) was obtained by the Russian scientist D. I. Mendeleev and is often called the Mendeleev-Clapeyron equation.

EXPERIMENTAL VERIFICATION OF THE EQUATION OF STATE. The validity of the equation of state in the form (3.5) can be demonstrated with the help of the set-up shown in Fig. 31.

A tightly closed corrugated vessel is connected to a manometer M that indicates the pressure inside the vessel. The volume of the vessel can be changed by turning a screw S. The volume can be measured with the help of a ruler R. The temperature of a gas in the vessel is equal to the temperature of the ambient and is measured by a thermometer.

---

[1] The French physicist B. P. Clapeyron (1799-1864) worked in Russia for ten years.
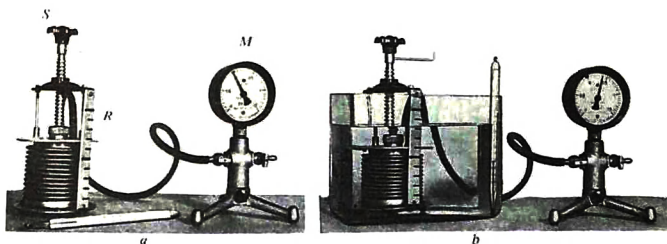
Fig. 31

By measuring the pressure p, temperature T and volume V of the gas in the initial state (Fig. 31a), we can calculate the quantity $p_1 V_1/T_1$ (it should be borne in mind that T is the absolute temperature and not the temperature on the Celsius scale).

Next we change the volume of the vessel and heat the gas by placing the vessel into hot water (Fig. 31b). Having measured again the gas pressure p_2, volume V_2 and temperature T_2, we can calculate the quantity $p_2 V_2/T_2$.

Within the accuracy limits of the given experimental set-up, the equation of state (3.5) is satisfied. Only when the pressure attains values of thousands of atmospheres, the deviations of experimental results from those predicted by the equation of state of an ideal gas become considerable. Dense gases cannot be treated as ideal even approximately.

3.2.   Gas Laws

Using the equation of state of an ideal gas, we can analyze the processes in which the mass and one of the three parameters (p, V, or T) remain unchanged. The quantitative relations between two parameters of a gas for a fixed value of the third parameter are called GAS LAWS.

Processes occurring at a constant value of one parameter are called ISOPROCESSES[1]. Isoprocesses are frequently encountered in nature and widely used in engineering.

ISOTHERMAL PROCESS. *The process of change in the state of a thermodynamic system at a constant temperature is called an isothermal process.* To maintain the constant temperature in a gas it must be able to exchange heat with a large system–a THERMOSTAT. It can be the atmospheric air provided that its temperature does not change noticeably during the process.

According to the equation of state (3.4) for an ideal gas, the product of

---

[1] From the Greek word *isos* meaning "equal".

Fig. 32
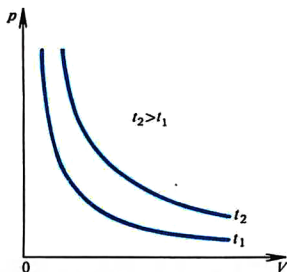
pressure and volume is the same for any state at a constant temperature:

$$pV = \text{const} \quad \text{for} \quad t = \text{const}. \tag{3.6}$$

If the temperature of a gas does not change, the product of the gas pressure and volume is constant for a given mass of the gas.

This law was established experimentally by the English scientist Boyle and somewhat later by the French physicist Mariotte. It is known in physics as BOYLE'S LAW.

Boyle's law is valid for any gas as well as for a gas mixture (like air). The deviations from this law become significant only for pressures exceeding the atmospheric pressure by about three orders of magnitude.

The validity of Boyle's law at pressures close to the atmospheric pressure can be demonstrated with the help of the set-up described in the previous section.

The dependence of the gas pressure on volume at constant temperature is represented graphically by a curve known as the ISOTHERM (Fig. 32). An isotherm of a gas shows the inverse proportionality between the pressure and the volume. Such a curve is called a hyperbola in mathematics.

Different isotherms correspond to different constant temperatures. As the temperature increases, the pressure also increases in accordance with the equation of state (3.4) if $V = \text{const}$. Therefore, the isotherm corresponding to a higher temperature t_2 lies above the isotherm corresponding to a lower temperature t_1.

ISOBARIC PROCESS. *The process of variation of the state of a thermodynamic system at a constant pressure is known as an isobaric process* [1].

According to Eq. (3.4), the ratio of the volume of a gas to its absolute temperature in any state remains constant at a constant pressure:

$$\frac{V}{T} = \text{const} \quad \text{at} \quad p = \text{const}. \tag{3.7}$$

---

[1] From the Greek word *baros* meaning "weight".
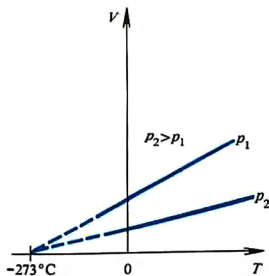
Fig. 33
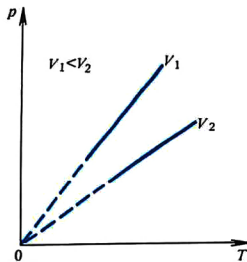
Fig. 34

For a fixed mass of a gas, the ratio of volume to temperature is constant if the gas pressure remains unchanged.

This law was established by the French scientist J. Charles and refined by Gay-Lussac in 1802 and is known as Charles' law. According to (3.7), the volume of a gas at a constant pressure depends linearly on temperature:

$$V = \text{const} \cdot T. \tag{3.8}$$

This dependence is graphically represented by a straight line called an ISOBAR (Fig. 33).

Different isobars correspond to different pressures. As the pressure of a gas increases, the volume decreases according to Boyle's law at constant temperature. Therefore, the isobar corresponding to a higher pressure p_2 lies below the isobar corresponding to a lower pressure p_1.

In the low-temperature region, all isobars of an ideal gas converge at a point $T = 0$. This does not mean, however, that the volume of a real gas actually vanishes. All gases liquefy when cooled to very low temperatures, and the equation of state (3.4) is inapplicable to liquids.

ISOCHORIC PROCESS. *A change in the state of a thermodynamic system at a constant volume is called an isochoric process* [1].

The equation of state (3.4) implies that the ratio of the gas pressure to temperature remains constant in any state of a gas, the volume being constant throughout:

$$p/T = \text{const} \quad \text{for} \quad V = \text{const.} \tag{3.9}$$

For a fixed mass of a gas, the ratio of pressure to temperature remains constant if the volume of the gas does not change.

This law was established by Charles in 1787 and refined by Gay-Lussac and is known as the Gay-Lussac law. According to (3.9), the pressure of a gas linearly depends on temperature at a constant volume:

$$p = \text{const} \cdot T. \tag{3.10}$$

---

[1] From the Greek word *chorema* meaning "capacity" (volume).

This dependence is graphically represented by a straight line known as an ISOCHORE (Fig. 34). Different isochores correspond to different volumes. According to Boyle's law, as the volume of a gas is increased at a constant temperature, its pressure drops. Therefore, the isochore corresponding to a larger volume V_2 lies below the isochore corresponding to a smaller volume V_1.

According to Eq. (3.10), all isochores emerge from the point $T = 0$. This means that the *pressure of an ideal gas is equal to zero at the absolute zero temperature.*

3.3. Application of Properties of Gases in Engineering

Gases have a number of properties that make them irreplaceable in many technical devices.

GAS AS A SHOCK-ABSORBER. High compressibility, low density, and the possibility to control pressure of a gas make it a perfect shock-absorber in many devices.

This is used in the construction of motocar and bicycle tyres. When a wheel comes across a protuberance, the air in the tyre is compressed, and the shock transmitted to the axles of the wheel is much softer (Fig. 35). If the tyre were rigid, the axles would jump to the height of the protuberance.

GAS AS A WORKING MEDIUM IN ENGINES. The high compressibility and strong dependence of pressure and volume on the temperature of a gas make it an irreplaceable working medium for compressed-air engines as well as for heat engines.

In the compressed-air engines, the gas performs work when expanded at a nearly constant pressure. Compressed air opens doors in buses and electric trains by exerting pressure on a piston. It drives the pistons of air brakes in railway cars and lorries. Air-operated hammers and other pneumatic instruments are driven by compressed air. Small jet engines operating on compressed helium are mounted on spacecrafts for manual orientation.

For internal combustion engines used in motorcars, tractors, aeroplanes

Fig. 35

Working stroke
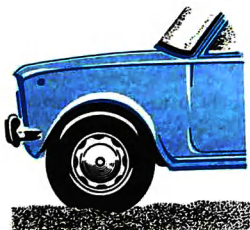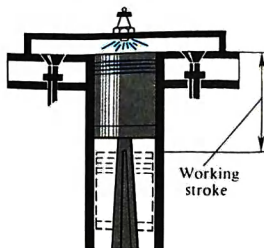
Fig. 36

and rockets, the working medium which drives the pistons or turbines is a gas at a high temperature. When a gas mixture is ignited in a cylinder, the temperature sharply increases to thousands of degrees, the pressure on the piston rises, and the expanding gas performs work for the duration of the working stroke of the piston (Fig. 36).

Gas is the only state of aggregation that can be used as a working medium in internal combustion engines. The heating of a liquid or a solid to such temperatures as those employed for gases would cause only a slight displacement of the piston.

Any weapon is essentially a heat engine. The force of pressure exerted by gaseous products of combustion of the explosives pushes a bullet from the barrel of a rifle or a shell from the mouth of a cannon. It is essential that this force does work over the entire length of the barrel. For this reason, the velocities of the bullet and shell reach values up to several hundreds of metres per second.

RAREFIED GASES. The ability of gases to expand unlimitedly complicates the process of obtaining gases at very low pressures, viz. in the state of vacuum. (In this state, gas molecules practically do not collide with one another and only strike the vessel walls.)

Ordinary piston pumps become ineffective because gas molecules can penetrate between the piston and the vessel walls. They cannot be used for obtaining pressures lower than a few tenths of one mm Hg. More complex vacuum pumps are used for creating high rarefactions. At present, the pressures are of the order of 10^{-10} Pa (10^{-12} mm Hg) have been attained.

High vacuum is mainly required in cathode-ray tubes and other electronic devices. Collisions between electrically charged particles (electrons) and gas molecules deteriorate the normal operation of these devices. Sometimes, vacuum has to be created in very large volumes as, for example, in elementary particle accelerators.

Vacuum is also required for manufacturing highly purified metals, for creating thermal insulation, and so on.

?
1. What equation is known as the equation of state?
2. Formulate the equation of state for an arbitrary mass of an ideal gas.
3. What is the value of the universal gas constant?
4. What is the relation between gas pressure and volume for an isothermal process?
5. Plot two isotherms for an ideal gas at temperatures T_1 and $T_2 > T_1$.
6. What is the relation between volume and temperature of a gas for an isobaric process?
7. What is the relation between pressure and temperature for an isochoric process?
8. How can isothermal, isobaric and isochoric processes be realized?
9. Give a qualitative explanation of gas laws on the basis of the molecular kinetic theory.
10. Why is the working medium in heat engines only in the gaseous form?

Problems with Solutions

The following situations are commonly encountered while solving problems concerning gas laws:

(a) The macroscopic parameters in the initial state and some parameters in the final state are known.

If one of the parameters does not change as the system goes over from the initial to the final state, we can use for isothermal processes Boyle's law in the form (3.6) or in the equivalent form $p_1/p_2 = V_2/V_1$ following from it. For isobaric processes, use should be made of Charles' law in the form (3.7) or in the equivalent form $V_1/V_2 = T_1/T_2$. For isochoric processes, we use Gay-Lussac's law in the form (3.9) or in the equivalent form $p_1/p_2 = T_1/T_2$.

If all the three parameters change, the equation of state (3.4) or (3.5) should be used.

(b) A part of macroscopic parameters for a gas in a certain state is known, and the remaining parameters should be found. In such cases, the Mendeleev-Clapeyron equation (3.4) is the most convenient.

(c) In many problems, the graphs representing various processes are required. To plot the graphs, we must know the relation between the state parameters. In the general case, this dependence is given by the equation of state and, in special cases, by gas laws.

While solving problems, one should have a clear idea about the initial state of a system and about a process that transforms it to the final state.

1. A cylinder having a volume $V_1 = 0.02$ m^3 and containing air under a pressure $p_1 = 4 \times 10^5$ Pa is connected with another cylinder having a volume $V_2 = 0.06$ m^3 and preliminarily evacuated. Find the pressure p established in the cylinders if the temperature is constant.

Solution. The air contained in the first cylinder will occupy the entire volume $V_1 + V_2$ of the two cylinders. According to Boyle's law, we have

$$\frac{p}{p_1} = \frac{V_1}{V_1 + V_2}.$$

Hence the required pressure is

$$p = \frac{p_1 V_1}{V_1 + V_2} = 10^5 \text{ Pa}.$$

2. By what fraction of the initial volume will the volume of a gas increase when heated from the initial temperature $t = 27\,°C$ by $1\,°C$ at a constant pressure?

Solution. Let V_1 and $T_1 = 273$ K $+ t = 300$ K be the initial values of the state parameters of the gas, and V_2 and $T_2 = T_1 + 1$ K be the final values of the same parameters.

According to Charles' law, we have

$$\frac{V_2}{V_1} = \frac{T_2}{T_1}.$$

The equality does not change if we subtract unity from the left- and right-

54
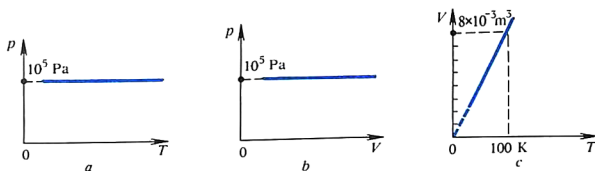
Fig. 37

hand sides of this equation:

$$\frac{V_2}{V_1} - 1 = \frac{T_2}{T_1} - 1.$$

This gives

$$\frac{V_2 - V_1}{V_1} = \frac{T_2 - T_1}{T_1}, \quad V_2 - V_1 = \frac{1}{300} V_1.$$

The gas volume has increased by 1/300 of its initial value.

3. Under normal conditions (i.e. at $t_0 = 0$ C and the normal atmospheric pressure $p_0 = 101\,325$ Pa), the density of air is $\rho = 1.29$ kg/m^3. Find the mean molar mass M of air.

Solution. The equation of state of an ideal gas under standard (normal) conditions has the form $p_0 V_0 = (m/M) R T_0$. Here $T_0 = 273$ K and $R = 8.31$ J/(K·mole). Hence

$$M = \frac{mRT_0}{V_0 p_0} = \frac{\rho_0 T_0}{p_0} R \simeq 0.029 \text{ kg/mole.}$$

4. Plot the isobars for 2 g of hydrogen under the normal atmospheric pressure p_0 in the p-T, p-V and V-T coordinates.

Solution. In the p-T and p-V coordinates, isobars are straight lines parallel to the T- and V-axis respectively (Fig. 37a and b).

Since $V = (m/M)(R/p_0) T$, the V vs. T graph is a straight line passing through the origin. Considering that $m = 0.002$ kg, $M = 0.002$ kg/mole, $R = 8.31$ J/(mole·K) and $p_0 \simeq 10^5$ Pa, we can write $V = BT$, where $B = (m/M)(R/p_0) \simeq 8 \times 10^{-5}$ m^3/K. In particular, for $T = 100$ K, $V \simeq 8 \times 10^{-3}$ m^3. The V vs. T graph is shown in Fig. 37c.

Exercise 3
1. A gas is compressed isothermally from a volume $V_1 = 8$ l to $V_2 = 6$ l. As a result, the gas pressure increases by $\Delta p = 4$ kPa. What was the initial pressure p_1 of the gas?
2. A compressor used for driving mechanical picks sucks in $V = 100$ l of air per second from the atmosphere. What is the number of mechanical picks that can be fed by this compressor if each pick requires $V_i = 100$ cm^3 of air per second under a pressure $p = 5$ MPa? The atmospheric pressure $p = 100$ kPa.

Fig. 38
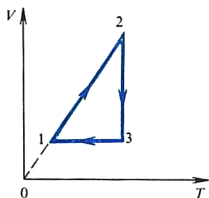
3. Plot the isotherms for 2 g of hydrogen at 0 C in the p-V, V-T, and p-T coordinates.

4. Determine the temperature of a gas contained in a closed vessel if upon heating by 1 K its pressure increases by 0.4% of its initial value.

5. As a certain mass of a gas goes over from one state to another, its pressure decreases and the temperature rises. How does the volume of this mass of the gas change?

6. Determine the volume of a mole of an ideal gas under standard conditions.

7. Find the mass of air in your classroom at 20 C and at normal atmospheric pressure, assuming that the molar mass of air is 0.029 kg/mole.

8. A certain gaseous substance has a density of 2.5 kg/m³ at a temperature of 10 °C and at normal atmospheric pressure. Find the molar mass of this substance.

9. A cylinder whose volume is 0.03 m³ contains a gas under a pressure of 1.35×10^6 Pa at a temperature of 455 °C. What would be the volume of this gas under normal conditions ($t_0 = 0$ C and $p_0 = 101\,325$ Pa)?

10. The height of the Lenin Peak in the Pamir mountains is 7134 m. At this height, the atmospheric pressure is 3.8×10^4 Pa. What is the air density at the top of this mountain at 0 C if its value is 1.29 kg/m³ under normal conditions?

11. Figure 38 shows a graph illustrating the change in the state of an ideal gas in the V-T coordinates. Represent the same process in the p-V and p-T coordinates.

12. Express the root-mean-square velocity of a molecule in terms of the universal gas constant and the molar mass.

13. A gas is contained in a cylinder at a temperature of 15 C. Find the ratio of the initial and the final pressures of the gas when 40% of the gas flows out of the cylinder and the temperature drops thereby by 8 °C.

Main Points of Chapter 3

The relation existing between thermodynamic parameters is given by the equation of state. All sufficiently rarefied gases (ideal gases) obey the Mendeleev-Clapeyron equation of state:

$$pV = \frac{m}{M} RT,$$

56

where p is the pressure, V is the volume, m is the mass, M is the molar mass, T is the absolute temperature and $R = 8.31 \ J/(K \cdot mole)$ is the universal gas constant.

The equation of state represents, as particular cases, the gas laws relating the change of two thermodynamic parameters at a constant value of the third parameter:

For a given mass of a gas, $pV = const$ for $T = const$ (Boyle's law)

$V/T = const$ for $p = const$ (Charles' law);

$p/T = const$ for $V = const$ (Gay-Lussac's law).

4 THE FIRST LAW OF THERMODYNAMICS

4.1. Internal Energy

It was mentioned in the course of "Junior Physics" that every macroscopic body possesses an internal energy.

The concept of internal energy of macroscopic bodies plays a very important role in the analysis of thermal phenomena. This is associated with the fundamental law of nature, viz. the LAW OF CONSERVATION OF ENERGY.

The discovery of the law of energy conservation became possible only when it was proved that in addition to the mechanical energy, *macroscopic bodies also possess an internal energy which is contained in them. This energy participates in the overall balance of energy transformations in nature.*

When a puck sliding over an ice surface stops under the action of friction, its mechanical (kinetic) energy does not vanish but is transferred to the molecules of ice and puck, which are in random motion. Uneven surfaces of bodies rubbing against each other are deformed in motion, and the intensity of random motion increases. The two bodies are heated, which indicates an increase in their internal energy.

The reverse transition of the internal energy into the mechanical energy can also be easily observed. If we heat water in a test tube closed with a cork, the internal energy of water increases. Water boils, and the steam pressure increases to such an extent that the cork is knocked out of the test tube and flies away. The kinetic energy of the cork increases at the expense of the internal energy of the steam. The expanding steam performs work and is cooled. Its kinetic energy decreases thereby.

From the point of view of the molecular kinetic theory, the internal energy of a macroscopic body is equal to the sum of the kinetic energies of random motion of all the molecules (or atoms) relative to the centre of mass of the body and the potential energies of interaction among all the molecules (but not with the molecules of other bodies)[1]. It is practically impossible to

---

[1] The internal energy also includes the energy of motion and interaction of particles constituting atoms and molecules. However, this energy is constant if there are no significant variations of temperature.

calculate the internal energy of a body (or its change) by taking into account the motion of individual molecules and their relative positions because of the huge number of molecules in macroscopic bodies. Therefore, we must determine the mean value of the internal energy (or its change) in terms of macroscopic parameters which can be measured directly.

INTERNAL ENERGY OF AN IDEAL MONATOMIC GAS. A monatomic gas consisting of individual atoms and not molecules is a system with the simplest properties. Inert gases, such as helium, neon and argon, are examples of monatomic gases. Let us calculate the internal energy of an ideal monatomic gas.

The potential energy of the molecules constituting an ideal gas is assumed to be zero since they do not interact with one another, except for the brief durations of their collisions. *The internal energy of an ideal gas is the kinetic energy of thermal motion of its molecules.*

In order to calculate the internal energy of a mass m of an ideal monatomic gas, we must multiply the mean kinetic energy of an atom, $\bar{E} = (3/2)kT$, by the number of atoms $N = (m/M) \times N_A$. Considering that $kN_A = = R$, we obtain the internal energy of the ideal gas:

$$U = \frac{3}{2}\frac{m}{M}RT. \tag{4.1}$$

The internal energy of an ideal monatomic gas is proportional to its absolute temperature.

This energy does not depend on the volume and other macroscopic parameters of a system. A change in the internal energy of a given mass of an ideal gas takes place only when its temperature changes:

$$\Delta U = \frac{3}{2}\frac{m}{M}R\,\Delta T.$$

If the mass of a gas increases, the internal energy also increases ($U \propto m$). The internal energy depends on the type of gas, i.e. on its molar mass: $U \propto 1/M$. This is due to the fact that the larger the value of M, the smaller the number of atoms in a given mass of the gas.

If an ideal gas consists of more complex molecules than a monatomic gas, its internal energy is also proportional to the absolute temperature, but the proportionality factor between U and T is different. This is explained by the fact that in addition to the translatory motion complex molecules also perform rotational motion. The internal energy of such gases is equal to the sum of the energies of translational and rotational motions of molecules.

DEPENDENCE OF INTERNAL ENERGY ON MACROSCOPIC PARAMETERS. Thus, we have established that the internal energy of an ideal gas is determined by a single parameter, viz. temperature. It does not depend on the volume since the potential energy of interaction of its molecules is assumed to be zero.

For real gases, liquids and solids, the mean potential energy of interaction of molecules differs from zero. True, it is much lower than the mean kinetic energy for gases, but for liquids and solids these energies are comparable.

The mean potential energy of molecular interaction depends on the volume of a substance since a change in the volume leads to a change in the mean distance between the molecules. Consequently, the internal energy generally depends on the volume V as well as on the temperature T.

The values of macroscopic parameters T, V, etc. uniquely determine the state of a body. Therefore, these parameters also determine the internal energy of macroscopic bodies.

The internal energy U of a macroscopic body is uniquely determined by the parameters which characterize the state of the body, viz. by its temperature and volume.

Let us now consider processes which may lead to a change in the internal energy. It was mentioned in the course of "Junior Physics" that there are two kinds of such processes: those in which work is done and those in which heat is transferred. We shall first consider in greater detail the concept of work in thermodynamics.

4.2. Work in Thermodynamics

WORK IN MECHANICS AND THERMODYNAMICS. Work is defined in mechanics as the product of the magnitudes of the force and displacement, multiplied by the cosine of the angle between these vectors. Work is done when a force acts on a moving body and is equal to the change in its kinetic energy.

In thermodynamics, we do not consider the motion of a body as a whole and analyze the displacement of various parts of a macroscopic body relative to one another. As a result, the volume of the body changes but its velocity remains zero. Like in mechanics, work is defined by the change in the energy of the body, but now we speak about the change in the internal energy and not in the kinetic energy.

THE CHANGE IN THE INTERNAL ENERGY WHEN WORK IS DONE. Why does the internal energy of a body change upon compression or extension? Why, in particular, is air heated when we inflate a bicycle tyre?

The cause of the change in temperature during the compression of a gas is as follows. *In elastic collisions between gas molecules and a moving piston, their kinetic energy changes.* During the head-on collisions with the molecules, the piston imparts to them a fraction of its mechanical energy, and as a result the gas is heated. The piston behaves like a football player, who kicks the flying ball with his foot and imparts to the ball a velocity which is much higher than the velocity of the ball before the blow.

Conversely, if a gas expands, the velocities of its molecules decrease upon collisions with the escaping piston, and as a result the gas is cooled. The football player acts in the same way when he wants to reduce the velocity of a flying ball or to stop it: his foot moves away from the ball as if giving way to the ball.

Compression or expansion also changes the mean potential energy of molecular interaction since the mean distance between the molecules changes.
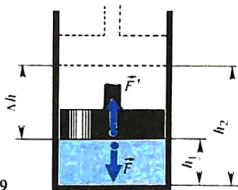
Fig. 39
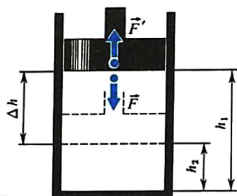
Fig. 40

CALCULATION OF WORK. Let us consider as an example a gas contained under a piston in a cylinder (Fig. 39) and calculate the work associated with a change in the volume. It is easier to calculate first the work done by the force $\vec{F}'$ exerted by the gas on the piston rather than the work done by the external force $\vec{F}$ exerted on the gas by the external body (piston). According to Newton's third law, $\vec{F}' = -\vec{F}$.

The magnitude of the force exerted by the gas on the piston is $F' = pS$, where p is the gas pressure and S is the area of the piston. Let us suppose that the gas expands and the piston is displaced in the direction of the force $\vec{F}'$ by a small distance $\Delta h = h_2 - h_1$. If the displacement is small, the gas pressure can be treated as constant.

The work done by the gas is

$$A' = F' \Delta h = pS(h_2 - h_1) = p(Sh_2 - Sh_1). \tag{4.2}$$

This work can be expressed in terms of the change in the gas volume. The initial volume $V_1 = Sh_1$, while the final volume is $V_2 = Sh_2$. Therefore,

$$A' = p(V_2 - V_1) = p\,\Delta V, \tag{4.3}$$

where $\Delta V = V_2 - V_1$ is the change in the gas volume.

The expanding gas does a positive work since the direction of the force coincides with the direction of the piston's displacement. In the process of expansion, the gas transfers an energy to the surrounding bodies.

If a gas is compressed, formula (4.3) for the work done by the gas remains valid, but now $V_2 < V_1$ and hence $A' < 0$ (Fig. 40).

The work A done by external forces on a gas differs from the work A' only in sign: $A = -A'$, since the force $\vec{F}$ acting on the gas is directed against the force $\vec{F}'$, while the displacement is the same. Thus, the work done by external forces on a gas is given by

$$A = -A' = -p\,\Delta V. \tag{4.4}$$

The minus sign indicates that the work done by an external force during compression is positive when $\Delta V = V_2 - V_1 < 0$. In this case $A > 0$ because the external force coincides in direction with the displacement. When external forces do a positive work on a gas, they impart to it a certain energy. On the contrary, the work of external forces is negative during an
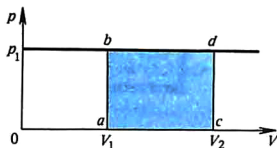
Fig. 41
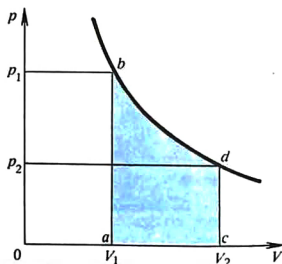
Fig. 42

expansion of a gas $(A < 0)$ since $\Delta V = V_2 - V_1 > 0$. The force and displacement have now opposite directions.

Formulas (4.3) and (4.4) are valid not only for compression or expansion of a gas in a cylinder but also for a small change in the volume of any system. If a process is isobaric ($p = $ const), these formulas can be used for large variations of the volume as well.

GEOMETRICAL INTERPRETATION OF WORK. If the gas pressure is constant, the work A' done by a gas can be given a simple geometrical interpretation.

Let us plot the gas pressure as a function of the volume (Fig. 41). Here the area of the rectangle $abdc$ bounded by the graph $p_1 = $ const, the V-axis and the segments ab and cd (each of which is equal to the pressure) is numerically equal to the work done by the gas (4.3):

$$A' = p_1 (V_2 - V_1) = |ab||cd|.$$

In the general case, however, the gas pressure is changed when the volume of a gas changes arbitrarily. For example, it decreases in inverse proportion to the volume for an isothermal process (Fig. 42). In this case, we must divide the total change in the volume into small fractions, calculate the elementary (small) work for each fraction, and then find their sum. As before, the work is numerically equal to the area of the figure bounded by the p-V graph, the V-axis and the segments ab and cd equal to the pressures p_1 and p_2 in the initial and final states.

?   1. Which physical quantities determine the internal energy of a body?
2. Give examples of mutual transformations of the mechanical energy and the internal energy in engineering and in everyday life.
3. What is the expression for the internal energy of an ideal monatomic gas?
4. Compare the internal energies of a mole of hydrogen and a mole of helium at the same temperature.
5. Why is a gas heated upon compression?
6. Write the expression for the work done by external forces during compression and expansion of a body.

4.3. Amount of Heat[1]

The internal energy of a gas in a cylinder can be changed not only by doing work but also by heating the gas (Fig. 43). If we fix the piston, the volume of the gas remains unchanged, but its temperature, and hence the internal energy, increase.

The process of energy transfer from one body to another without doing work is known as heat exchange or heat transfer.

The energy supplied to a body as a result of heat transfer is called the amount of heat. This term is also applied to the energy given away by a body as a result of heat transfer.

MOLECULAR PATTERN OF HEAT TRANSFER. During heat transfer, an interaction between the slow molecules of a cold body and fast molecules of a hot body takes place at the interface between two bodies in contact. As a result, the kinetic energies of the molecules level out, the velocities of molecules of the cold body increase, and those of the molecules of the hot body become lower.

During heat transfer, there is no energy transformation: a fraction of the internal energy of the hot body is transferred to the cold body.

AMOUNT OF HEAT AND HEAT CAPACITY. It is known from the course of "Junior Physics" that in order to heat a body of mass m from temperature t_1 to temperature t_2, the amount of heat that should be supplied to the body is

$$Q = cm(t_2 - t_1) = cm\,\Delta t. \qquad (4.5)$$

When the body is cooled, its final temperature t_2 is lower than the initial temperature t_1, and the amount of heat given away by the body is negative.

The coefficient c in formula (4.5) is known as the SPECIFIC HEAT. The specific heat is the amount of heat received or given away by 1 kg of a substance upon a change in its temperature by 1 K.

Specific heat is expressed in joules per kilogram per kelvin. Different bodies require different amounts of heat for increasing their temperature by 1 K. For example, the specific heat for water is 4190 J/(kg·K), while for copper it is 380 J/(kg·K).

The specific heat depends not only on the properties of the substance but also on the process in which heat is transferred. If a gas is heated at a constant pressure, it expands and does a work. A larger amount of heat is required to heat a gas by 1 °C at a constant pressure than to heat it by 1 °C at constant volume.

Liquids and solids expand insignificantly upon heating and their specific heats at constant pressure and at constant volume differ only slightly.

SPECIFIC LATENT HEAT OF VAPORIZATION. In order to vaporize a liquid, a certain amount of heat must be transferred to it. In this

---

[1] In this section, we briefly recollect the information on the amount of heat contained in the course of "Junior Physics".
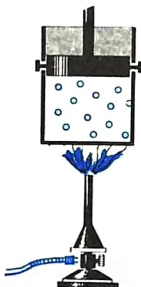
Fig. 43

transformation, the temperature of the liquid remains unchanged. The conversion of a liquid into vapour at a constant temperature does not lead to an increase in the kinetic energy of molecules but is accompanied by an increase in their potential energy, since the mean distance between gas molecules is many times longer than between liquid molecules.

The amount of heat required for vaporizing 1 kg of a liquid at constant temperature is called the specific latent heat of vaporization.

denoted by r and expressed in joules per kilogram.

The specific latent heat of vaporization of water is very large: $r_w = 2.256 \times 10^6$ J/kg at 100 °C. The specific latent heats of vaporization for other liquids (such as alcohol, ether, mercury or kerosene) are 3-10 times lower.

To vaporize a mass m of a liquid, the following amount of heat is required:

$$Q_v = rm. \qquad (4.6)$$

During condensation of vapour, the same amount of heat is liberated:

$$Q_c = -rm. \qquad (4.7)$$

SPECIFIC LATENT HEAT OF FUSION. During the melting of a crystalline solid, the entire heat supplied to it is spent to increase the potential energy of molecules. The kinetic energy of molecules remains unchanged since melting (fusion) occurs at a constant temperature.

The amount of heat λ required for melting 1 kg of a crystalline substance at its melting point is known as the specific latent heat of fusion.

During the crystallization of 1 kg of the substance, the same amount of heat is liberated. The specific latent heat of fusion for ice is quite large: $\lambda_{ice} = 3.34 \times 10^5$ J/kg.

In order to melt a crystalline body of mass m, the following amount of heat is required:

$$Q_f = \lambda m. \qquad (4.8)$$

64

The amount of heat liberated upon the crystallization of a body is given by

$$Q_{cr} = -\lambda m. \tag{4.9}$$

1. What is the amount of heat?
2. What does the specific heat of a substance depend on?
3. Give the definition of the specific latent heat of vaporization.
4. What is the specific latent heat of fusion?
5. When is the amount of heat negative?

4.4. The First Law of Thermodynamics

ENERGY CONSERVATION LAW. By the middle of the 19th century, it had been proved in many experiments that the mechanical energy never vanishes without any trace. For example, when a hammer strikes a piece of lead, the latter is heated to a certain temperature after the blow.

Many similar observations and generalizations of experimental results formed the basis for the LAW OF CONSERVATION OF ENERGY.

Energy in nature cannot be created from outside nor can it vanish: the total amount of energy remains constant, it is only converted from one form to another.

The law of energy conservation governs all phenomena in nature and links them together. It is fulfilled without any exception (there is no example in which this law is violated).

This law was discovered in the middle of the 19th century by the German scientist J. R. Mayer (1814-1878), and by the English scientist J. P. Joule (1818-1889). In its most complete form, this law was formulated in the works of the German scientist H. L. Helmholtz (1821-1894).

THE FIRST LAW OF THERMODYNAMICS. *The law of energy conservation and transformation, applied to thermal phenomena is known as the first law of thermodynamics.*

In thermodynamics, we deal with bodies whose centre of mass remains virtually fixed. The mechanical energy of bodies is constant, and only their internal energy may change.

Till now, we have considered the processes in which the internal energy of a system changes either as a result of doing work or due to heat transfer to the surrounding bodies. In the general case, as a system goes over from one state to another, its internal energy changes both as a result of doing work and due to heat transfer. The first law of thermodynamics is formulated just for such general cases:

The change in the internal energy of a system going over from one state to another is equal to the sum of the work done by external forces and the amount of heat supplied to the system:

$$\Delta U = A + Q. \tag{4.10}$$

In the particular case of a closed system, no work is done on it ($A = 0$),

Fig. 44

Fig. 45 Thermal insulation
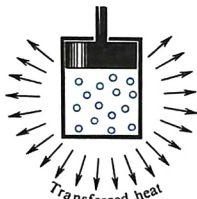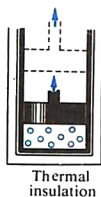
and there is no heat exchange with the surroundings ($Q = 0$). According to the first law of thermodynamics, in this case $\Delta U = U_2 - U_1 = 0$, or $U_1 = U_2$. *The internal energy of a closed system remains unchanged* (is conserved).

Instead of the work A done by external forces on a system, the work A' done by the system on external bodies is often considered. If we recall that $A = -A'$ (see Sec. 4.2), the first law of thermodynamics (4.10) can be written as follows:

$$Q = \Delta U + A'. \tag{4.11}$$

The amount of heat supplied to a system is spent in increasing its internal energy and in the work done by the system on surrounding bodies.

THE IMPOSSIBILITY OF CREATING A PERPETUAL MOTION MACHINE. The first law of thermodynamics implies that it is impossible to construct a perpetual motion machine, viz. a device capable of doing an unlimited work without consuming a fuel or any other material. If no heat is supplied to a system ($Q = 0$), then according to (4.11), the work A' can be done only at the expense of a decrease in the internal energy: $A' = -\Delta U$. After the energy reserve has been exhausted, the machine stops operating.

WORK AND AMOUNT OF HEAT AS CHARACTERISTICS OF THE PROCESS OF THE CHANGE IN ENERGY. A system in a given state always has a certain internal energy. *It would be wrong to say*, however, *that a certain amount of heat or work is contained in it*. Work, as well as the amount of heat, is a quantity characterizing the CHANGE IN THE ENERGY of the system as a result of some process.

The internal energy of a system may change by the same amount as a result of work done by the system or due to a heat transfer to the surroundings. The energy of a heated gas in a cylinder can be decreased by cooling it without doing work (Fig. 44). On the other hand, it can lose the same amount of energy by moving the piston without transferring heat to the surroundings. For this the cylinder walls and the piston must be thermally insulated (Fig. 45).

In this course of "Senior Physics", we shall deal with various forms of energy, the ways of their transformation and transfer.

4.5. Application of the First Law of Thermodynamics to Various Processes

On the basis of the first law of thermodynamics, we can draw important conclusions about the course of a process. Let us consider various processes in which one of physical parameters remains unchanged (isoprocesses). Suppose that a system is an ideal gas (this is the simplest case).

ISOCHORIC PROCESS. In this process, the volume does not change, and therefore the work of the gas is zero. According to Eq. (4.11), the change in the internal energy is equal to the amount of supplied heat:

$$\Delta U = Q. \tag{4.12}$$

If the gas is heated, then $Q > 0$, and $\Delta U > 0$ as well. This means that the internal energy of the gas increases. For the case of cooling, we have $Q < 0$ and $\Delta U = U_2 - U_1 < 0$. The change in the internal energy is negative, and hence the internal energy of the gas decreases.

ISOTHERMAL PROCESS. In this process ($T =$ const), the internal energy of an ideal gas given by Eq. (4.1) remains unchanged. According to formula (4.11), the entire heat supplied to the system is spent on doing work:

$$Q = A'. \tag{4.13}$$

If the gas receives heat ($Q > 0$), it expands and performs positive work ($A' > 0$). Conversely, if the gas gives heat to the surroundings (to a thermostat), then $Q < 0$ and $A' < 0$. The work of external forces on the gas is positive in this case.

ISOBARIC PROCESS. According to formula (4.11) the amount of heat supplied to the gas is spent in this process for increasing the system's internal energy and for doing work at a constant pressure.

If the gas is heated ($Q > 0$), it expands and does positive work ($A' > 0$). As a result, its internal energy increases ($\Delta U > 0$).

In isobaric cooling ($Q < 0$), the gas contracts, external forces do positive work on it ($A > 0$), and its internal energy decreases ($\Delta U < 0$).

ADIABATIC PROCESS. The processes considered above were described in Chapter 3 before the first law of thermodynamics was introduced. Let us consider now a process occurring in a system that does not exchange heat with the surroundings. *Processes occurring in a thermally insulated system are called adiabatic processes.*

In such a process, $Q = 0$, and according to Eq. (4.10), the change in the internal energy is only associated with the work done on the system:

$$\Delta U = A. \tag{4.14}$$

Naturally, we cannot enclose the system in an envelope that would exclude heat transfer. However, in some cases we can assume that real processes are very close to adiabatic processes. For this they must be sufficiently rapid so that no noticeable heat transfer takes place between the system and the surroundings during the process.

According to Eq. (4.14), when positive work is done on a system as, for

5*

Fig. 46

Atomizer
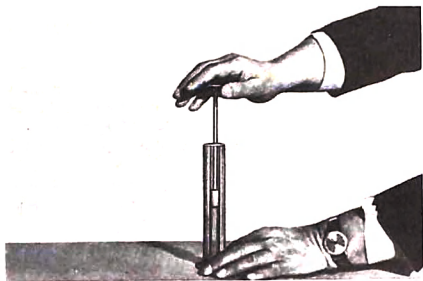
Fig. 47

example, during compression of a gas, the internal energy of the gas increases. This means that the gas temperature rises. Conversely, an expanding gas performs positive work $(A' > 0)$, and its internal energy decreases, i.e. the gas is cooled.

Heating of a gas upon rapid compression can be demonstrated with the help of a transparent cylinder with a tightly fitting piston (Fig. 46). If we put a piece of cotton wetted by ether on the bottom of the cylinder and quickly lower the piston, ether vapour ignites.

Heating of air upon a rapid compression is used in Diesel engines. These engines have no ignition system like the one used for the air-fuel mixture in internal combustion petrol engines. In Diesel engines atmospheric air and not air-fuel mixture is sucked in a cylinder. By the end of the compression stroke, liquid fuel is introduced into the cylinder with the help of a special atomizer (Fig. 47). By this time, the air temperature in the cylinder is so high that the fuel ignites.

During the operation of high-power air compressors, the air temperature increases to such an extent that a special system of cooling the cylinders has to be employed. Adiabatic cooling of gases upon expansion is used in devices for liquefaction of gases.

Cooling of gases during an adiabatic expansion occurs on a gigantic scale in the Earth's atmosphere. Heated air goes up and expands since the atmospheric pressure decreases with increasing height. This expansion is accompanied by a significant cooling. As a result, water vapour condenses and clouds are formed.

HEAT TRANSFER IN A CLOSED SYSTEM. Let us consider heat transfer in a system consisting of several bodies which initially have different temperatures, for example, the heat transfer between water in a vessel and a hot iron ball immersed in water. We shall assume that the system is insulated from the surroundings well enough and that its internal energy does not change (the system is closed). No work is done in the system. Consequently, according to the first law of thermodynamics (4.10), the

change in the energy of any body in the system is equal to the amount of heat given away or received by this body until thermal equilibrium sets in the system ($\Delta U'_i = Q_i$). Adding similar expressions for all bodies of the system and considering that the total internal energy remains unchanged ($\Delta U_1 + \Delta U_2 + \Delta U_3 + \ldots = 0$), we obtain the following equation:

$$Q_1 + Q_2 + Q_3 + \ldots = 0. \tag{4.15}$$

This equation is known as the HEAT BALANCE EQUATION. Here Q_1, Q_2, Q_3, ... are the amounts of heat received or given away by the bodies. These amounts of heat are given by formula (4.5) or formulas (4.6)-(4.9) if heat transfer involves transformations between the states of aggregation (a substance goes from the liquid to gaseous or solid state or vice versa).

The heat balance equation was first established experimentally when heat exchange between bodies was observed in a calorimeter, viz. a device insulating bodies from the surroundings to the maximum possible extent. The construction of the calorimeter was described in the course of "Junior Physics".

1. Formulate the first law of thermodynamics.
2. When is a change in the internal energy negative?
3. Compare the work done by a gas in an isothermal expansion from volume V_1 to volume V_2 and during an isobaric expansion from V_1 to V_2.
4. Why is it wrong to say that a system can store a certain amount of heat or work?
5. Which process is called an adiabatic process?
6. Figure 48 shows pressure versus volume curves for adiabatic and isothermal processes. Why is the adiabat more steep?
7. Write the heat balance equation.

4.6. Irreversibility of Processes in Nature

The law of energy conservation states that the amount of energy in a closed system remains unchanged in all transformations. However, this law does not indicate which energy transformations are possible. Moreover, many processes allowed by the law of conservation of energy never occur in actual practice.
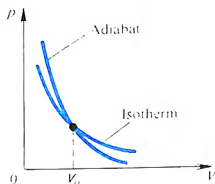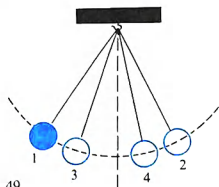

Fig. 48

Fig. 49

EXAMPLES OF IRREVERSIBLE PROCESSES. Heated bodies are gradually cooled, transferring their energy to cold surrounding bodies. The reverse process of heat transfer from a cold body to a hot one does not contradict the law of energy conservation, but such a process has been never observed.

Another example. The oscillations of a pendulum displaced from the equilibrium position attenuate (Fig. 49; *1, 2, 3,* and *4* are consecutive positions of the pendulum). The mechanical energy of the pendulum decreases at the expense of the work of the forces of friction, and the temperature of the pendulum and surrounding air (and hence their internal energy) slightly increase. The reverse process, i.e. an increase in the amplitude of oscillations of the pendulum due to cooling of the pendulum and surrounding air is also permissible from the point of view of energy. This process, however, is never observed. The mechanical energy can be spontaneously transformed into internal energy, but not the other way round. The ordered motion of a body as a whole is replaced by a random thermal motion of the molecules constituting it.

GENERAL CONCLUSION ABOUT IRREVERSIBILITY OF PROCESSES IN NATURE. Heat transfer from a hot to a cold body and the transformation of the mechanical energy into the internal energy are the examples of the most typical irreversible processes. The number of such examples is practically unlimited. All these examples indicate that processes in nature occur in a certain direction, which is not reflected in the first law of thermodynamics at all. *All macroscopic processes in nature occur only in one direction and cannot spontaneously proceed in the reverse direction.* All processes in nature are irreversible, the most tragic among them being the ageing and death of living organisms.

EXACT FORMULATION OF AN IRREVERSIBLE PROCESS. In order to understand correctly the essence of the irreversibility of processes in nature, the following ammendment should be made. A process is called irreversible if the reverse process to it may occur only as a link of a more complex process. For example, the amplitude of pendulum oscillations can be increased by pushing the pendulum by hand. This increase, however, does not emerge by itself but becomes possible as a result of a more complex process including the motion of the hand.

It is possible in principle to transfer heat from a cold body to a hot one. But for this purpose we need a refrigerator which consumes energy.

"REVERSE" MOVIE. A brilliant illustration of the irreversibility of phenomena in nature is running a motion picture in the reverse direction. A jump in water, for example, looks as follows in this case. Still water in a swimming pool starts to swirl, there appear legs leaping up, and finally, the entire body of the diver emerges from water. The surface of water soon becomes still. The velocity of the diver decreases, and we see him standing on the springboard. The process on the screen could be observed in reality if the processes were reversible. The processes on the screen seem unreal to us because we are accustomed to a certain course of processes, and the impossibility of their reversal is beyond doubt. However, a process like raising up the diver to the springboard does not contradict the energy

conservation law, or the laws of mechanics, or some other law except THE SECOND LAW OF THERMODYNAMICS.

THE SECOND LAW OF THERMODYNAMICS. This law indicates the direction of possible energy transformations and thus expresses the irreversibility of processes in nature. It was established as a direct generalization of experimental results.

There are several formulations of the second law. Despite their apparent difference, they essentially express the same thing and hence are equivalent.

The German scientist R. J. Clausius formulated this law as follows: it is impossible to transfer heat from a cold system to a hot one in the absence of other simultaneous changes in the two systems or in the surroundings.

Here, the experimental fact of a certain course in heat transfer is stated: heat is always transferred spontaneously from a hot body to a cold body. Of course, in refrigerators heat is transferred from a cold body to a hot one, but this transfer is accompanied by "other changes in the surrounding bodies": cooling is realized at the expense of work.

The importance of this law lies in that conclusion about the irreversibility of other processes in nature besides heat transfer can be drawn from it. If heat could be transferred spontaneously from cold to hot bodies, this would make it possible to reverse other processes also.

4.7. Principle of Operation and Efficiency of Heat Engines

The internal energy stored in the Earth's crust and in the oceans can be assumed to be practically unlimited. However, to have energy in store is not enough. It is necessary to know how to use this energy for driving machines in factories and plants, transport facilities, tractors and other machines, to rotate the rotors of generators, and so on. Mankind needs engines, i.e. DEVICES CAPABLE OF DOING WORK. Most of the engines used nowadays are HEAT ENGINES, i.e. DEVICES THAT CONVERT THE INTERNAL ENERGY OF FUEL INTO THE MECHANICAL ENERGY.

PRINCIPLES OF OPERATION OF HEAT ENGINES. In order to make an engine do work, a pressure difference must be created on the two sides of its piston or turbine blades. In all types of heat engines, this pressure drop is attained by increasing the temperature of the working medium by hundreds or thousands of degrees over the temperature of the surroundings. This increase in temperature is attained by burning fuel.

The working medium of all heat engines is a gas (See Sec. 3.3). A gas does work during its expansion. We denote the initial temperature of the working medium (gas) by T_1. In steam turbines and engines, steam acquires this temperature in a boiler. In internal combustion engines and gas turbines, the combustion of a fuel takes place inside the engine. THE TEMPERATURE T_1 IS KNOWN AS THE HEATER TEMPERATURE (see coloured plate 1).

THE ROLE OF THE COOLER. Having performed work, a gas loses its energy and inevitably cools down to a certain temperature T_2. This temperature cannot be lower than the temperature of the ambient since

otherwise the gas pressure would be lower than the atmospheric pressure, and the engine would be incapable of doing work. In actual practice, the temperature T_2 is somewhat higher than the temperature of the ambient. It is called THE COOLER TEMPERATURE. The cooler is the atmosphere or a special equipment (called CONDENSER) for cooling and condensation of waste steam.

Thus, the working medium expanding in an engine cannot spend its entire internal energy in doing work. A fraction of heat is inevitably transferred to a cooler (atmosphere) together with the waste steam or exhaust gas of internal combustion engines and gas turbines. This fraction of internal energy is lost. The heat engine does work at the expense of the internal energy of the working medium. In this process, heat is transferred from a hot body (heater) to a cold body (cooler).

The schematic diagram of a heat engine is shown in the coloured plate 1. The working medium of the engine receives an amount of heat Q_1 as a result of combustion of a fuel, does work A' and transfers an amount of heat $Q_2 < Q_1$ to the cooler.

THE EFFICIENCY OF A HEAT ENGINE. The impossibility of complete conversion of internal energy into work stems from the irreversibility of processes in nature. If heat could spontaneously return from cooler to heater, the internal energy would be completely converted into work with the help of any heat engine.

According to the energy conservation law, the work done by an engine is

$$A' = |Q_1| - |Q_2|, \qquad (4.16)$$

where Q_1 is the amount of heat received from the heater and Q_2 is the heat transferred to the cooler.

The efficiency of a heat engine is the ratio of the work A' done by it to the amount of heat received from the heater [1]:

$$\eta = \frac{A'}{|Q_1|} = \frac{|Q_1| - |Q_2|}{|Q_1|} = 1 - \frac{|Q_2|}{|Q_1|}. \qquad (4.17)$$

Since a certain amount of heat is transferred to the cooler in all engines, we always have $\eta < 1$.

The efficiency of a heat engine is determined by the difference between the temperatures of the heater and cooler. It must be proportional to the difference $T_1 - T_2$ between the temperatures of the heater and cooler since it cannot operate when $T_1 - T_2 = 0$.

THE MAXIMUM EFFICIENCY OF HEAT ENGINES. The laws of thermodynamics allow us to calculate the maximum possible efficiency of a heat engine operating at a temperature T_1 of the heater and a temperature T_2 of the cooler. This estimate was made for the first time by the French engineer and scientist Sadi Carnot (1796-1832) in his work "Réflexions sur la puissance motrice du feu" (1824).

---

[1] In steam turbines, the boiler plays the role of the heater while in internal combustion engines the fuel combustion products are the "heaters". (*Transl. note.*)

Carnot invented an ideal heat engine with an ideal gas as the working medium. He obtained the following expression for the efficiency of this engine:

$$\eta_{max} = \frac{T_1 - T_2}{T_1}. \tag{4.18}$$

As expected, the efficiency of the Carnot engine is directly proportional to the difference in the absolute temperatures of the heater and cooler and inversely proportional to the absolute temperature of the heater. The latter circumstance is due to the fact that the amount of heat Q_1, as well as the internal energy, is proportional to the absolute temperature of the heater.

The principal significance of this formula is that, as was proved by Carnot, *any real heat engine operating at heater temperature T_1 and the cooler temperature T_2 cannot have an efficiency exceeding the efficiency of the corresponding ideal heat engine.*

Formula (4.18) sets a theoretical limit for the maximum efficiency of heat engines. It shows that the efficiency of a heat engine is the higher, the higher the temperature of the heater and the lower the temperature of the cooler. The efficiency η can be equal to unity only when the temperature of the cooler is equal to the absolute zero.

However, the cooler temperature practically cannot be lower than the ambient temperature. Naturally, we can increase the temperature of the heater, but any (solid) material has a limited thermal stability, or high-temperature strength. During heating, it gradually loses its elastic properties, and when the temperature becomes sufficiently high, it just melts.

The main efforts of engineers are concentrated now on increasing the efficiency of engines by reducing friction between the parts, losses due to incomplete combustion of fuel, and so on. In this direction, there do exist considerable possibilities for increasing the efficiency. For example, the initial and final temperature of steam in steam turbines are about $T_1 = 800$ K and $T_2 = 300$ K. The maximum efficiency for these temperatures is

$$\eta_{max} = \frac{T_1 - T_2}{T_1} \simeq 0.62, \quad \eta_{max} \simeq 62\%.$$

However, the actual value of the efficiency is $\eta \simeq 40\%$ because of various types of energy losses. The maximum efficiency of internal combustion engines is about 44%. To increase the efficiency of heat engines and to bring it closer to the maximum possible value is a very important problem in engineering.

4.8. Role of Heat Engine. Heat Engines and Protection of Environment

ROLE OF HEAT ENGINES. Heat engines (mainly high-power steam turbines) are widely used in thermoelectric plants where they drive the rotors of electric current generators. More than

80% of the electric power in the USSR is generated at thermoelectric plants.

Heat engines (steam turbines) are installed in nuclear power plants. High-temperature steam is generated in these plants at the expense of the energy liberated in the fission of atomic nuclei.

Heat engines are also used in all types of modern vehicles. Automobiles are equipped with piston internal combustion engines in which air-fuel mixture is prepared in a carburettor outside the cylinder (carburettor engines), as well as those in which the air-fuel mixture is formed directly in the cylinder (Diesel engines). The latter engines are also widely used in tractors employed in agriculture.

Till the middle of this century, the railway transport mainly operated on steam engine. Nowadays, locomotives with Diesel and electric engines are frequently employed. However, electric locomotives in the long run consume energy produced by heat engines of thermoelectric plants.

In ships and steamboats both internal combustion engines and high-power steam turbines are used.

In aviation piston engines are installed on light aeroplanes, while turbojet and jet engines are used on huge liners (the latter engines are also heat engines). Jet engines are used on spacecraft as well.

Modern life cannot be imagined without heat engines. Without them, we could not have cheap electric power or any type of high-speed vehicles.

HEAT ENGINES AND PROTECTION OF ENVIRONMENT. The wide application of heat engines for producing energy in a form convenient to consumers affects environment to a greater extent then any other technological process.

In accordance with the laws of thermodynamics, electric or mechanical energy cannot be produced in principle without releasing considerable amounts of heat to the ambient. This inevitably leads to a gradual increase in the average temperature on the Earth. The consumed power amounts at present to 10^{10} kW. When this figure reaches 3×10^{12} kW, the average temperature will increase noticeably (by one degree). A further increase in temperature may lead to thawing of glaciers and a catastrophic increase in the level of the world ocean.

But this does not exhaust the negative consequences of utilization of heat engines. The furnaces of thermoelectric plants, the internal combustion engines of motorcars, etc. continuously discharge to the atmosphere substances which are harmful for plants and living organisms: sulphur compounds (formed during the combustion of coal), nitrogen oxides, hydrocarbons, carbon monoxide (CO) and so on. Especially harmful in this respect are motorcars, whose number is continuously increasing, while the purification of waste gases remains imperfect. In nuclear power plants, the disposal of reactive wastes is an important problem.

Moreover, the application of steam turbines in electric power plants requires large areas to be used as tanks for cooling waste steam. As the power of electric stations increases water demand rises in proportion. In 1980, 200 km^3 of water was consumed in the Soviet Union for this purpose, which amounts to 35% of water supply in all branches of national economy.

All this poses a number of serious problems for human society. Along with

the efforts to increase the efficiency of heat engines, measures should be taken to protect the environment. For this, the efficiency of devices preventing from the discharge of harmful substances to the atmosphere should be increased. Efforts should be made to attain a more complete combustion of fuel in automobile engines. Motorcars having too high amounts of CO in the waste gas should not be allowed on the roads. The projects of electromobiles competing with ordinary engines are being developed, and the possibility to use a fuel with no harmful substances in the waste gas, such as a hydrogen-oxygen mixture, is being discussed.

In order to save ground and water resources, it is expedient to construct complexes of electric power plants, and in the first place nuclear power plants, with a closed water supply cycle.

Another trend concerns an efficient utilization of energy and its economic consumption.

The solution of these problems is of vital importance for the mankind. Planned development of national economy can help solve these problems successfully. Environmental control, however, requires efforts on a world-wide scale.

?
1. Which processes are called irreversible?
2. Give examples of typical irreversible processes.
3. What other irreversible processes do you know?
4. Formulate the second law of thermodynamics.
5. Would it be a violation of the energy conservation law if rivers flew backwards?
6. Is the motion of the Moon around the Sun a reversible process?
7. What device is called a heat engine?
8. What role do the heater, the cooler and the working medium play in a heat engine?
9. Why cannot the internal energy of the ocean be used as a source of energy in heat engines?
10. Define the efficiency of a heat engine.
11. How can the maximum efficiency of a heat engine be determined?

Problems with Solutions

In problems to this chapter, the first law of thermodynamics is used in the form (4.10) or (4.11).

In solving problems, formulas (4.4) should be used for calculating work, and (4.5)-(4.9) for calculating heat. It must be borne in mind that the values of A, Q and ΔU can be either positive or negative.

Instead of the general formula for the first law of thermodynamics most problems employ formulas for special cases applied to definite processes. The problems concerning heat transfer in a closed system should be solved with the help of the heat balance equation (4.15).

As in the solution of problems on gas laws, in the problems to this section the initial and final states of a system should be clearly specified together with the parameters characterizing these states.

1. An aerostat having a volume $V = 500$ m^3 is filled with helium under a pressure $p = 10^5$ Pa. As a result of heating by solar radiation, the temperature of the gas in the aerostat increases from $t_1 = 10$ C to $t_2 = 25$ C. What is the increase in the internal energy of the gas?

Solution. Helium is a monatomic gas whose internal energy is determined by formula (4.1). This energy is $U_1 = (3/2)(m/M)RT_1$ at temperature T_1 and $U_2 = (3/2)(m/M)RT_2$ at T_2. The change in the energy is given by

$$\Delta U = U_2 - U_1 = \frac{3}{2}\frac{m}{M}R(T_2 - T_1).$$

The mass of helium is unknown, but it can be expressed with the help of the Mendeleev-Clapeyron equation in terms of the initial temperature, pressure and volume of the gas: $mR/M = pV/T_1$. Substituting the value of mR/M into the equation for the change in the internal energy, we obtain

$$\Delta U = \frac{3}{2}pV\left(\frac{T_2}{T_1} - 1\right) \simeq 4 \times 10^6 \text{ J}.$$

2. Carbon dioxide ($M = 0.044$ kg/mole) of mass $m = 0.20$ kg is contained in a cylinder under a heavy piston. The gas is heated by $\Delta T = 88$ K. Find the work done by the gas upon heating.

Solution. The gas expands under a constant pressure p created by the atmosphere and the piston. The work done by the gas is then

$$A' = p(V_2 - V_1),$$

where V_1 and V_2 are the initial and final volumes of the gas. Using the equation of state $pV = (m/M)RT$ for an ideal gas, we can express the products pV_1 and pV_2 through $(m/M)RT_1$ and $(m/M)RT_2$. This gives

$$A' = \frac{m}{M}R(T_2 - T_1) \simeq 3.3 \text{ J}.$$

3. A gas expanding in a cylinder having a cross-sectional area $S = 200$ cm^2 receives an amount of heat $Q = 1.5 \times 10^5$ J so that the gas pressure remains constant and is equal to $p = 2 \times 10^7$ Pa. What is the change in the internal energy of the gas, if the piston in the cylinder has moved by $\Delta h = 30$ cm?

Solution. According to the first law of thermodynamics (4.11), we have

$$Q = \Delta U + A',$$

where $A' = p\Delta V = pS\Delta h$ is the work done by the gas. Hence

$$\Delta U = Q - pS\Delta h = 30 \text{ kJ}.$$

4. Water of mass $m_2 = 0.30$ kg at a temperature $t_2 = 80$ C is poured in a calorimeter containing water of mass $m_1 = 0.20$ kg at a temperature $t_1 = 20$ C. The temperature of water in the calorimeter becomes $t = 50$ C. What is the heat capacity of the calorimeter? (The heat capacity of a body is the product of its mass and the specific heat: $C = cm$.)

Solution. According to the heat balance equation, the sum of the amounts of heat exchanged between bodies in a calorimeter is zero. Water at

temperature t_2 transfers an amount of heat $cm_2(t - t_2)$ while water at temperature t_1 receives an amount of heat $cm_1(t - t_1)$. The calorimeter receives the heat $C(t - t_1)$ since its initial temperature is the same as that of water in it t_1. Consequently, we can write

$$cm_2(t - t_2) + cm_1(t - t_1) + C(t - t_1) = 0,$$

whence

$$C = -\frac{cm_1(t_1 - t) + cm_2(t_2 - t)}{t - t_1} \simeq 420 \text{ J/k}.$$

5. Water drops constitute 90% of the mass of waste steam at a temperature $t_1 = 100$ C. To cool the steam, it is mixed with the same mass of cold water ($t_2 = 10$ C). What is the resulting temperature of water?

Solution. We denote the mass of cold water by m. Then the mass of steam is $0.1m$, while the mass of water drops is $0.9m$. The amount of heat given away during the condensation of steam and during the cooling of water formed is negative:

$$Q_1 = -0.1rm + 0.1cm(t - t_1).$$

The amount of heat given away by cooling drops is also negative:

$$Q_2 = 0.9cm(t - t_1), \text{ where } t < t_1.$$

The amount of heat received by the cold water is $Q_3 = cm(t - t_2)$. We write the heat balance equation:

$$Q_1 + Q_2 + Q_3 = 0,$$

$$-0.1rm + 0.1cm(t - t_1) + 0.9cm(t - t_1) + cm(t - t_2) = 0.$$

Hence

$$t = \frac{0.1r + c(t_1 + t_2)}{2c} \simeq 82\,^\circ\text{C}.$$

Exercise 4

1. What is the change in the internal energy of a monatomic ideal gas whose pressure increases by a factor of three as its volume decreases to half the initial value?
2. A gas having a pressure $p = 10^5$ Pa expands isobarically, doing work $A = 25$ J. What is the increase in the volume of the gas?
3. A thermodynamic system has received 200 J of heat. What is the change in the internal energy of the system if the work done by it is 400 J?
4. The rod of a pneumatic pick is driven by compressed air. The mass of air in the cylinder changes from 0.1 to 0.5 g during a piston stroke. Assuming that the air pressure and temperature ($27\,^\circ$C) in the cylinder are constant, determine the work done by the gas during a stroke. The molar mass of air is $M = 0.029$ kg/mole.
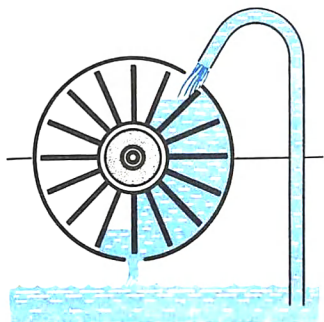5. Two tightly closed vessels having a volume of 1 l each are heated by

Fig. 50

identical burners. One vessel contains water and the other contains air. Which vessel will be heated first to 50 C? Why?

6. The following project for a perpetual motion machine is proposed (Fig. 50). A closed vessel is divided into two parts by a hermetic partition through which pass a tube and a water turbine in a jacket with two holes. The air pressure in the lower part is higher than in the upper part. Water rises in the tube and fills the open chamber. The next portion of water in the lower part is poured out of the turbine chamber where the level has reached the hole in the jacket. Why cannot this machine operate perpetually?

7. Is the work done by the gas in the processes *1-2, 2-3* and *3-1* shown in Fig. 38 positive or negative? Does the gas receive or give away heat in each of these processes?

8. The temperature of a gas whose mass is m and the molar mass is M is increased by ΔT first at a constant pressure p and then at a constant volume V. What is the difference in the amounts of heat supplied to the gas in these two cases?

9. What amount of heat is required for isochoric heating of 4 kg of helium by 100 K?

10. A gas performs 20 J work during its isothermal expansion. What amount of heat is supplied to the gas?

11. Calculate the increase in the internal energy of 2 kg of hydrogen as a result of its isobaric heating by 10 K. The specific heat of hydrogen at constant pressure is 14 kJ/(kg·K).

12. Four moles of an ideal monatomic gas are compressed in the cylinder of a compressor. Calculate the increase in the gas temperature after a piston stroke if the work of 500 J is done. The process should be treated as adiabatic.

13. Water vapour having a mass of 10 g and a temperature of 100 C is introduced into a calorimeter containing 0.25 kg of water at 25 C. Determine the temperature established in the calorimeter if its heat capacity is 1000 J/K.

14. A calorimeter contains 0.4 kg of water at 10 C. A piece of ice whose mass is 0.6 kg and temperature -40 C is put in water. What is the

temperature established in the calorimeter if its heat capacity is negligibly small?

15. What must be the temperature of a heater for obtaining in principle an 80"„ efficiency of the heat engine if the temperature of the cooler is 27 C?

16. The working medium of a heat engine receives from the heater $Q_1 = 1.5 \times 10^6$ J of heat and gives away $Q_2 = -1.2 \times 10^6$ J of heat to the cooler during a certain period of operation. Calculate the efficiency of the engine and compare it with the maximum possible efficiency if the temperatures of the heater and cooler are 250 C and 30 C.

Main Points of Chapter 4

Macroscopic bodies possess internal energy equal to the sum of the kinetic energies of random motion of all the molecules in the body and the potential energies of interaction of all the molecules with one another. The internal energy is uniquely determined by the thermodynamic parameters (temperature and volume). For an ideal monatomic gas, the internal energy depends only on its temperature:

$$U = \frac{3}{2} \frac{m}{M} RT.$$

According to the first law of thermodynamics, the change in the internal energy of a system going over from one state to another is equal to the sum of the work done by external forces on the system and the amount of heat received by it:

$$\Delta U = A + Q.$$

The work done on a thermodynamic system is $A = -p\,\Delta V$, where p is the pressure and ΔV is the change in the volume. The work done by the system in this case is $A' = -A = p\,\Delta V$. The heat transferred during heating or cooling is $Q = mc\,\Delta T$, where c is the specific heat and ΔT is the change in the temperature. Besides, heat is absorbed during vaporization and fusion and is liberated in condensation and crystallization (see formulas (4.6)-(4.9)).

Work and amount of heat are the characteristics of processes in which energy changes.

In an isochoric process ($V = $ const), the work is zero, and $\Delta U = Q$.

In an isothermal process ($T = $ const), the internal energy of an ideal gas remains unchanged, and $Q = A'$.

In an isobaric process ($p = $ const), the heat supplied to a system is spent for increasing the internal energy of the system and for doing work: $Q = \Delta U + A'$.

In an adiabatic process (in a thermally insulated system), $Q = 0$ and $\Delta U = A$.

If heat transfer takes place in a closed system without any work being done, the heat balance equation is valid:

$$Q_1 + Q_2 + Q_3 + \ldots = 0,$$

where $Q_1, Q_2, Q_3, \ldots$ are the amounts of heat received or given away by the bodies of the system.

Processes occurring in macroscopic bodies are irreversible. Typical irreversible processes are the transfer of heat from a hot to a cold body (and not vice versa) and the spontaneous transformation of mechanical energy into internal energy.

The generalization of experimental results concerning the irreversibility of processes led to the formulation of the second law of thermodynamics. It follows from the laws of thermodynamics that a heat engine can do work only in the process of heat transfer from a heater to a cooler. The maximum possible value of the efficiency of a heat engine is

$$\eta = \frac{T_1 - T_2}{T_1},$$

where T_1 is the temperature of the heater and T_2 is the temperature of the cooler.

To increase the efficiency of heat engines to the highest possible value is one of the most important problems in engineering.

5 MUTUAL CONVERSION OF LIQUIDS AND GASES

5.1. Saturated Vapour

The molecular kinetic theory not only explains why a substance can be in the gaseous, liquid or solid state but also clarifies the transitions of a substance from one state of aggregation to another.

EVAPORATION AND CONDENSATION. If water or some other liquid is kept in an open vessel, its amount gradually decreases. In this case, the EVAPORATION of the liquid takes place (the mechanism of this process was described in the course of "Junior Physics"). During random motion, some of the liquid molecules acquire such a high kinetic energy that they leave the liquid, having overcome the forces of attraction exerted by other molecules.

In addition to evaporation, the reverse process, viz. the return of a fraction of randomly moving vapour molecules to the liquid, also takes place. This process is known as CONDENSATION. If the vessel is open, the molecules of a liquid escaping from it may not return to the liquid. In such case, evaporation is not balanced by condensation, and the amount of liquid decreases. Liquids evaporate at a higher rate if the air flow above the liquid carries away the vapour formed, since the probability of the vapour molecules returning to the liquid becomes lower.

SATURATED VAPOUR. If a vessel containing a liquid is tightly closed, the amount of liquid soon stops decreasing. At a constant temperature, the liquid-vapour system attains the state of thermal equilibrium and remains in it as long as desired. Condensation occurs simultaneously with evaporation, the two processes balancing each other.

At the first moment after the liquid has been poured in a vessel and the latter has been closed, the liquid evaporates, and the density of vapour above it increases. However, the number of molecules returning to the liquid simultaneously increases. The higher the density of vapour, the larger the number of vapour molecules returning to the liquid. As a result, a dynamic equilibrium is ultimately established between the liquid and its vapour in the closed vessel at a constant temperature. The number of molecules escaping from the surface of the liquid is on the average equal to the number of

81

vapour molecules returning to the liquid during the same time.

A vapour which is in a dynamic equilibrium with its liquid is called the saturated vapour. This term emphasizes that a larger amount of vapour cannot be contained in a given volume at a given temperature.

If we pump the air out of a vessel containing a liquid, only saturated vapour will be contained above the surface of the liquid.

SATURATED VAPOUR PRESSURE. What will happen to a saturated vapour if the volume occupied by it is reduced (this can be done, for example, by compressing a vapour in equilibrium with its liquid in a cylinder under the piston so that the temperature of the content remains unchanged)?

As we compress the vapour, the equilibrium is violated. At the initial moment, the vapour density will slightly increase, and a larger number of molecules will go over from gas into liquid than from liquid to gas. This process occurs until the equilibrium sets in again, and the density, and hence the number density of molecules acquire the initial values. Consequently, the *molecular concentration of a saturated vapour at constant temperature does not depend on the volume.*

Since the pressure is proportional to the number density of molecules according to the formula $p = nkT$, the independence of the concentration (number density) of a saturated vapour on the volume implies that the saturated vapour pressure does not depend on the volume occupied by the vapour.

The vapour pressure p_0 which is independent of the volume and at which the liquid is in equilibrium with its vapour is called the saturated vapour pressure.

If a saturated vapour is compressed, an increasingly large fraction of its molecules goes over to the liquid state. A given mass of liquid occupies a smaller volume than its vapour of the same mass. As a result, the volume of vapour decreases, its density remaining unchanged.

In this discussion we have used more than once the terms "gas" and "vapour". A gas and a vapour do not differ in principle, and these terms are in general equivalent. However, we normally deal with a certain (relatively narrow) temperature interval of the ambient. The term "gas" is normally applied to the substances for which the saturated vapour pressure is higher than the atmospheric pressure under ordinary temperatures (for example, carbon dioxide). On the contrary, we use the term "vapour" when the saturated vapour pressure at room temperature is lower than the atmospheric pressure and the substance is more stable in the liquid state (for example, water vapour).
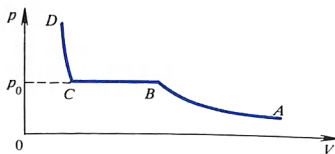

Fig. 51

82

The independence of the saturated vapour pressure on the volume has been established in numerous experiments on isothermal compression of a vapour which is in equilibrium with its liquid. Suppose that a large volume of a substance is in the gaseous state. In the course of isothermal compression, its density and pressure increase (segment AB of the isotherm in Fig. 51). As the pressure p_0 is attained, vapour starts to condense. During the further compression of the saturated vapour, its pressure does not change until the entire vapour condenses into liquid (segment BC of the straight line in Fig. 51). Then the pressure increases abruptly upon compression (segment CD of the curve) since the compressibility of liquids is low.

The curve shown in Fig. 51 is called the ISOTHERM OF A REAL GAS.

5.2. Temperature Dependence of Saturated Vapour Pressure. Boiling. Critical Temperature

TEMPERATURE DEPENDENCE OF SATURATED VAPOUR PRESSURE. The state of a saturated vapour can be approximately described by the equation of state (3.4) for an ideal gas, its pressure being approximately defined by the formula

$$p_0 = nkT. \tag{5.1}$$

Pressure increases with temperature. Since the *saturated vapour pressure does not depend on volume, it is determined only by temperature.*

However, the experimentally obtained dependence $p_0(T)$ is not linear as that for an ideal gas at a constant volume. As the temperature increases, the saturated vapour pressure increases at a higher rate than the pressure of an ideal gas (Fig. 52, segment AB).

The reason behind this is as follows. When a liquid and its vapour is heated in a closed vessel, a fraction of the liquid is evaporated. As a result, according to formula (5.1), the *vapour pressure increases not only due to an increase in temperature, but also due to any increase in the molecular concentration (density) of vapour.* The increase in pressure with temperature is mainly determined just by the increase in the concentration. The principal difference in the behaviour of an ideal gas and saturated vapour consists in that the mass of vapour in a closed vessel changes with the temperature (or with a change in volume at a constant temperature). The liquid partially evaporates or, on the contrary, the vapour is partially condensed. Nothing of this sort occurs in an ideal gas.

After the entire liquid has evaporated, its vapour ceases to be saturated upon further heating, and its pressure increases linearly with the absolute temperature at constant volume (segment BC in Fig. 52).

BOILING. As the temperature increases, the evaporation of the liquid becomes more intense. Finally, the liquid starts to boil. *During boiling rapidly growing vapour bubbles are formed in the entire volume of the liquid and rise to the surface.* The temperature of the liquid remains constant during this process.

Let us determine the conditions under which boiling begins. A liquid always contains dissolved gases which can be seen at the bottom, on the
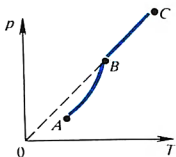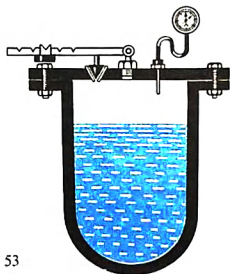
Fig. 52

Fig. 53

walls of the vessel and on dust particles suspended in the liquid. Vapour inside the bubbles is saturated. As the temperature rises, the saturation vapour pressure increases, and the bubbles increase in size. Under the action of buoyancy, they move to the surface. If the upper layers of the liquid have a lower temperature, the vapour in the bubbles condenses in these layers. The pressure abruptly drops, and the bubbles collapse. The collapse is so rapid that the bubble walls produce a sort of a miniexplosion. A large number of such explosions produce a noise characterizing the process of ebullition. When the entire liquid is uniformly heated, the collapse of bubbles ceases, and the bubbles come to the surface. The liquid starts to boil. If you watch a kettle on the stove, you will notice that it stops making noise before boiling begins.

The temperature dependence of the saturated vapour pressure explains why the boiling point of a liquid depends on the pressure on its surface. A vapour bubble may grow when the saturated vapour pressure inside it slightly exceeds the pressure in the liquid, which is the sum of the air pressure on its surface (external pressure) and the hydrostatic pressure of the liquid column.

Boiling begins at a temperature at which the saturated vapour pressure in the bubbles becomes equal to the pressure in the liquid.

The higher the external pressure, the higher the boiling point. For example, if the pressure in a boiler reaches 1.6×10^6 Pa, the water in it does not boil even at $200\,°C$. The sterilization of medical instruments in autoclaves (tightly closed vessels) is carried out under an elevated pressure (Fig. 53).

On the contrary, by lowering pressure, we make the boiling point of a liquid lower. We can make water boil at room temperature by pumping out air and water vapour from a flask containing water (Fig. 54). In the mountains, the atmospheric pressure decreases with increasing height, and hence the boiling point of water becomes lower. At a height of 7134 m (the Lenin peak in Pamir), the pressure is about 4×10^4 Pa (300 mm Hg), and the boiling point of water is about $70\,°C$. Under such conditions, it is impossible, for example, to boil an egg.
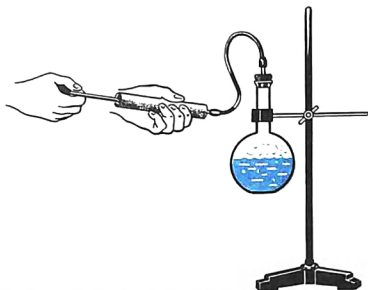
84

Fig. 54

The difference in boiling points of liquids is due to different values of their saturated vapour pressures. The higher the saturated vapour pressure, the lower the boiling point of a liquid since the saturated vapour pressure becomes equal to the atmospheric pressure at a lower temperature. For example, the saturated vapour pressure of water at $100\,°C$ is $101\,325$ Pa (760 mm Hg), while for mercury vapour, it is only 117 Pa (0.88 mm Hg). Mercury boils at $357\,°C$ under normal pressure.

SUPERHEATED LIQUID. It is possible to obtain a liquid at a temperature above the boiling point for a given pressure. (Such a liquid is called SUPERHEATED.) For this purpose, the liquid should be thoroughly purified and degassed. In such a liquid, the centres of evaporation, i.e. small air bubbles, are virtually absent. A superheated liquid is in an unstable state. It is sufficient to add chalk powder to it, which contains much air in the pores, to make it boil so furiously that it flows out of the vessel. In this case, the temperature of the liquid drops to the boiling point, since evaporation of the liquid requires a certain amount of energy.

CRITICAL TEMPERATURE. The density of saturated vapour increases with temperature as well as with pressure. On the contrary, the density of a liquid in equilibrium with its vapour decreases with heating due to the expansion of the liquid. If we plot in the same figure the temperature dependences of the densities of a liquid and its vapour, the curve will be inclined downwards for the liquid and upwards for its vapour (Fig. 55).

At a certain temperature called the CRITICAL temperature, the two curves merge, i.e. the density of the liquid becomes equal to the density of the vapour.

The temperature at which the difference in the physical properties of a liquid and its saturated vapour vanishes is known as the critical temperature.

At this temperature, the density (and pressure) of a saturated vapour attains its maximum value, while the density of the liquid which is in equilibrium with the vapour becomes minimal. The specific latent heat of vaporization decreases with increasing temperature and becomes zero at the critical temperature.
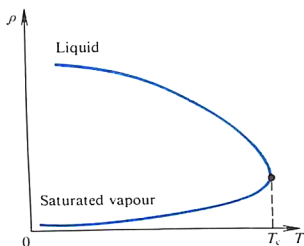
85

Fig. 55

Each substance is characterized by its own critical temperature. For example, the critical temperature for water is $\simeq 375$ C and for the liquid carbon dioxide (IV) $CO_2 \simeq 31$ C.

The main feature of critical temperature is that a gas cannot be liquefied under any pressure at temperatures above the critical.

5.3.   Humidity of Air

WATER VAPOUR IN THE ATMOSPHERE. Despite enormously large surfaces of oceans, seas, lakes and rivers, water vapour in the atmosphere is not saturated. Displacement of air masses leads to evaporation dominating over condensation in some regions of our planet at a given moment, while in other regions condensation prevails. However, atmospheric air practically always contains a certain amount of water vapour.

The content of water vapour in air, or, its humidity, is characterized by a number of quantities.

PARTIAL PRESSURE OF WATER VAPOUR. Atmospheric air is a mixture of various gases and water vapour. Each gas makes its contribution to the total pressure exerted by the air on bodies in it. The pressure exerted by the water vapour in the absence of all other gases is called the partial pressure of water vapour. The partial pressure of water vapour is one of the characteristics of air humidity. It is expressed in the units of pressure (pascals) or in millimetres of mercury column.

RELATIVE HUMIDITY. The partial pressure of water vapour alone cannot be used to judge the degree of saturation of a water vapour under given conditions. However, the degree of saturation determines the evaporation intensity and moisture losses for living organisms. For this reason, a quantity characterizing the degree of saturation of water vapour at a given temperature, viz. the RELATIVE HUMIDITY, is introduced.

The relative humidity φ of air is the percentage ratio of the partial pressure p of water vapour contained in air at a given temperature to the saturated vapour pressure p_0 at the same temperature:

$$\varphi = \frac{p}{p_0} 100\%. \qquad (5.2)$$

86

DEW POINT. If water vapour is cooled under a constant pressure, it becomes saturated sooner or later. This can be proved by analyzing the temperature dependence of saturated vapour pressure (Fig. 56).

Suppose that p_1 is the partial pressure of water vapour at a temperature t_1. The state of the vapour is represented by point A. If the vapour is cooled to a temperature t_d at $p_1 = $ const, the vapour becomes saturated, and its state is now represented by point B. *The temperature t_d at which water vapour becomes saturated is called the dew point.*

When air is cooled to the dew point, water vapour begins to condense (a mist appears and dew precipitates).

The dew point characterizes the humidity of air since it allows us to determine the partial pressure of water vapour and relative humidity. Indeed, if we know the dew point, we thus know the water vapour partial pressure p_1. It can be found with the help of the table containing the values of saturated vapour pressures corresponding to different temperatures (see the table on the end sheet). The saturated vapour pressure p_0 corresponding to the temperature t_1 is also taken from the table. From the values of p_1 and p_0, we can find the relative humidity $\varphi = (p_1/p_0) \, 100\%$.

HYGROMETERS.[1] Air humidity is measured with the help of special instruments: HYGROMETERS and PSYCHROMETERS[2] (wet-and-dry-bulb hygrometers).

Dew-point hygrometers make it possible to determine directly the dew point. A simple instrument of this type (Fig. 57) consists of a metallic box B with a well-polished front wall W, containing a readily evaporating liquid (ether). The air is passed through the box with the help of a rubber bulb, causing an intense evaporation of the ether and a rapid cooling of the box. A thermometer inserted into the box indicates the temperature at which dew drops precipitate on the polished surface of the wall. We can assume that the pressure in the region adjoining the wall is constant since this region is a part of the atmosphere, and a decrease in pressure is compensated by an increase in the vapour content. The appearance of dew indicates that the water vapour has become saturated. Knowing the temperature of the air and the dew point, we can find the partial pressure of water vapour and the relative humidity with the help of the table containing the values of saturated vapour pressure at various temperatures.

PSYCHROMETER. A psychrometer (wet-and-dry-bulb hygrometer) consists of two thermometers (Fig. 58). One of the thermometers has a dry bulb, and indicates the temperature of air. The bulb of the other thermometer is covered by a piece of wet cloth whose end is immersed in water. As water evaporates, the bulb of the thermometer is cooled. The higher the relative humidity, the less intense the evaporation and the higher the temperature indicated by the wet-bulb thermometer. At a relative humidity of 100%, water does not evaporate at all, and the two thermometers indicate the same temperature. Using special tables, we can

---

[1] From the Greek *hygros* meaning "wet".
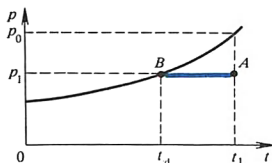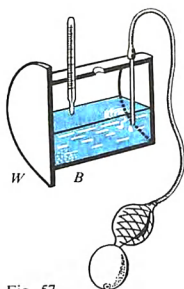[2] From the Greek *psychros* meaning "cold".
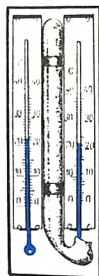
Fig. 56

Fig. 57

Fig. 58

determine the relative humidity of air from the difference in the temperatures of the dry-bulb and wet-bulb thermometers. Psychrometers make it possible to determine the humidity with a higher degree of accuracy than hygrometers.

THE ROLE OF HUMIDITY. The humidity of air determines the intensity of evaporation of moisture from the surface of the skin of a human body. In turn, the evaporation of moisture is important for maintaining a constant temperature of a human body. In spacecrafts, the most favourable relative humidity for human beings (40-60%) is maintained along with the right temperature and pressure.

The knowledge of humidity is essential in meteorology for weather forecasts. Although the content of water vapour in the atmosphere is relatively small (about 1%), its role in atmospheric phenomena is significant. The condensation of water vapour results in the formation of clouds, followed by the precipitation of rain or snow. This process is accompanied by the liberation of a large amount of heat. Conversely, evaporation of water involves an absorption of heat.

It is necessary to maintain a certain constant humidity in many technological processes (such as textile industry or confectionery production). The storage of books or pieces of art also requires the maintaining of air humidity at a certain level. On the walls of museums, you can see psychrometers.

?

1. Give the definition of saturated vapour.
2. Why is the saturated vapour pressure independent of volume?
3. Why does the saturated air pressure increases with temperature at a higher rate than the pressure of an ideal gas?
4. Why does boiling point of a liquid increase with pressure?
5. What is the critical temperature?
6. Define the relative humidity of air.
7. How can the relative humidity be determined from the known dew point?

Problems with Solutions

While solving problems, it should be borne in mind that the pressure and density of a saturated vapour do not depend on its volume and are determined only by the temperature. Away from the critical point, the equation of state for an ideal gas is approximately valid for describing a saturated vapour. However, the mass of saturated vapour does not remain constant during its compression or heating.

1. A closed vessel of volume $V_1 = 0.5$ m³ contains $m = 0.5$ kg of water. The vessel is heated to $t = 147$ C. By what amount must the vessel volume be increased so that it contains only the saturated vapour? The saturated vapour pressure at $t = 147$ C is $p_0 = 4.7 \times 10^5$ Pa.

Solution. Under a pressure p_0, the saturated vapour occupies the volume

$$V = \frac{mRT}{p_0 M} = 0.2 \text{ m}^3,$$

where $M = 0.018$ kg/mole is the molar mass of water. The vessel volume $V_1 > V$, and the vapour is not saturated. For it to become saturated, its volume should be decreased by

$$\Delta V = V_1 - V = V_1 - \frac{mRT}{p_0 M} = 0.3 \text{ m}^3.$$

2. The relative humidity of air in a closed vessel is $\varphi_1 = 84\%$ at a temperature $t_1 = 5$ C and $\varphi_2 = 30\%$ at a temperature $t_2 = 22$ C. Find the ratio of saturated vapour pressures at t_2 and t_1.

Solution. At $T_1 = 278$ K, the pressure of water vapour in the vessel is $p_1 = \varphi_1 p_{01}$, where p_{01} is the saturated vapour pressure at T_1. At $T_2 = 295$ K, $p_2 = \varphi_2 p_{02}$. Since the volume is constant, we have, in accordance with Gay-Lussac's law,

$$\frac{p_1}{p_2} = \frac{T_1}{T_2},$$

whence

$$\frac{p_{02}}{p_{01}} = \frac{\varphi_1 T_2}{\varphi_2 T_1} \simeq 3.$$

Exercise 5

1. Water has different levels in the limbs of a closed U-tube. Can we say that in addition to saturated water vapour, there is air above the water? Why?
2. In what way can we make water boil by cooling a vessel containing it?
3. What will be the change in the boiling point and in the latent heat of vaporization of water if the vessel containing it is lowered to the bottom of a deep well?

4. What is the density of vapour in the bubbles rising to the surface of water boiling at the atmospheric pressure?

5. There is a cold autumn rain outdoors. Wet washing is hung in the room. Will it be dried sooner if we open a window?

6. A pump delivers water to a height of 10.3 m. To what height can this pump lift water boiling at $100°C$, if its piston moves very slowly?

7. The dew point determined with the help of a dew-point hygrometer is $10°C$. Find the relative humidity of air if its temperature is $20°C$ (use the table on the end sheet of the textbook).

8. The relative humidity in a room of volume $V = 120$ m^3 is $\varphi = 60\%$, at $15°C$. Determine the mass of water vapour contained in the air if the saturated vapour pressure at this temperature is $p_0 = 12.8$ mm Hg.

9. The relative humidity φ_1 in a room is 20%, at $t = 20°C$. What mass of air has to be evaporated for increasing the relative humidity to $\varphi_2 = 50\%$, if the volume V of the room is 40 m^3? The density of saturated water vapour at this temperature is $\rho_0 = 1.73 \times 10^{-2}$ kg/m^3.

Main Points of Chapter 5

A liquid and its vapour can be in a dynamic equilibrium in which the number of molecules escaping from the liquid during a certain interval of time is equal to the number of molecules returning from the vapour to the liquid during the same time. A vapour in equilibrium with its liquid is called saturated. The saturated vapour pressure does not depend on the volume and is determined only by the temperature.

A liquid boils at a temperature for which the saturated vapour pressure in the bubbles becomes equal to the pressure in the liquid.

As the temperature increases, the density of saturated vapour rises, while the density of the liquid falls. At a certain temperature called the critical temperature, the difference in the physical properties of the liquid and its saturated vapour vanishes, and their densities equalize.

Atmospheric air is a mixture of various gases and water vapour. The content of water vapour in the air, or its humidity, is characterized by several quantities. The pressure exerted by water vapour in the absence of other gases is called the partial pressure of water vapour. The relative humidity of air is the percentage ratio of the partial pressure of water vapour contained in air at a given temperature to the saturated vapour pressure at the same temperature.

6 SURFACE TENSION OF LIQUIDS

6.1. Surface Tension

In addition to gravitation, elasticity and friction, there is one more force that acts in the world surrounding us but it is usually neglected. This force acts along the tangents to surfaces of all liquids.

Let us imagine a straight line drawn on the surface of a liquid and dividing this surface into two parts. The forces that pull these parts away from each other are known as the FORCES OF SURFACE TENSION. These forces are quite weak, but nevertheless they play a significant role in nature.

MECHANISM OF SURFACE TENSION. The easiest way to reveal the nature of surface tension is to observe the formation of a water drop at the spout of a loose tap. An attentive observer can see how the drop gradually increases in volume, a neck is formed and the drop is separated (Fig. 59). Not much imagination is required to see that the water is as if inclosed in an elastic sack which breaks away when its strength is insufficient to hold a large mass of water.

Of course, there is actually nothing but water in the drop, but the surface layer of water behaves as a stretched elastic film. The envelope of a soap bubble produces the same impression. It resembles the thin stretched rubber of a balloon.

Put carefully a sewing needle on the water surface. The surface film bends and does not let the needle sink. For the same reason, some light insects can rapidly slide over the water surface like skaters on ice (see coloured plate 2).

The bending of the film of water on the surface of a dense-mesh sieve prevents carefully poured water from flowing through the sieve. A cloth is a sort of a sieve formed by woven threads. Surface tension strongly hampers the penetration of water through the cloth. For this reason, cloth is not wetted through immediately.

If the force of gravity did not press a liquid to the bottom of a vessel or to other solid surface, the surface film, in its tendency to contract, would give the liquid a spherical shape. The smaller the drop, the more significant the surface forces in comparison with mass forces (gravitation). Therefore, small drops of dew have a nearly spherical shape (see coloured plate 2). During

Fig. 59

free fall, when the state of weightlessness sets in, the raindrops are almost spherical in shape.

In a spacecraft orbiting a planet under zero-gravity conditions, not only individual drops but also large masses of liquids have a spherical shape.

THE ORIGIN OF SURFACE TENSION. We can give a visual explanation to the emergence of surface tension. If a large group of individuals has a property to attract one another, or individuals are drawn towards one another by their own wish, the result is the same: they gather in a cluster like a swarm of bees. Each individual "strives" to get inside this cluster, and the surface of the cluster contracts, approaching the spherical surface. This is a model for the emergence of surface tension.

Molecules of water or some other liquid attract one another and strive to come closer. Each molecule on the surface is attracted by the remaining molecules located within the liquid, and hence has a tendency to sink. Since the liquid possesses fluidity because of molecular jumps from one "sedentary" position to another, it assumes a shape for which the number of molecules on its surface is minimal. The minimum surface at a given volume is that of the sphere. The area of the liquid surface is reduced, which is manifested as the surface tension.

These arguments reveal that the surface forces are of different origin in comparison with the elastic forces emerging in a stretched rubber film. The elastic force becomes weaker as the rubber contracts, while the surface tension does not change at all as the surface area of the film is reduced, since the density of the liquid (and hence the average intermolecular distance on the surface) does not change.

Thus, the emergence of surface tension cannot be explained as easily as the emergence of elastic forces which depend on the change in the intermolecular distance. Surface tension emerges in a more complicated rearrangement of the shape of the entire liquid, its volume remaining unchanged.

6.2. Force of Surface Tension

MEASUREMENT OF THE FORCE OF SURFACE TENSION. Let us suspend a Π-shaped wire with the length l of side AB from a spring balance (Fig. 60). The force of gravity acting on the wire is balanced by the upward elastic force $\vec{F}_0 = -m\vec{g}$. Then we

immerse the wire in a glass with a soap solution so that side AB is submerged. As we slowly lower the glass, a soap film is formed, and the spring of the balance stretches so that the pointer indicates a larger elastic force $\vec{F}_1$. This means that the soap film acts on side AB of the wire with a downward force.

The force of surface tension is the force acting along the surface of a liquid at right angles to the line bounding this surface and tending to contract it as much as possible.

Knowing the scale factor of the spring balance, we can easily determine this force. The film acts on the wire with the force $2\vec{F}$, where $\vec{F}$ is the surface tension exerted by one of the two surfaces of the film (Fig. 61). We can write the following equation for the forces in equilibrium:

$$\vec{F}_1 + m\vec{g} + 2\vec{F} = 0$$

or (for the magnitudes of the forces)

$$F_1 - mg - 2F = 0.$$

Considering that $mg = F_0$, we obtain

$$F = \frac{F_1 - mg}{2} = \frac{F_1 - F_0}{2}.$$

We can assume that the surface tension $\vec{F}$ is proportional to the length l of the surface layer of the liquid, i.e. to the length of the wire. Indeed, in all regions of the surface layer of the liquid, the molecules are in identical conditions, and the same force must act per unit length of the boundary of the surface. This assumption can be verified by taking another wire with side AB twice as long.

SURFACE TENSION. The ratio of the magnitude F of the force of surface tension acting on the boundary of the surface layer of length l to this length is a constant which does not depend on the length l. This quantity is known as the SURFACE TENSION and is denoted by σ:

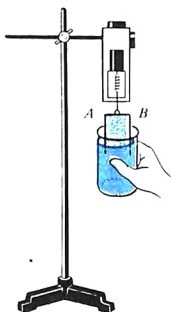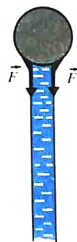$$\sigma = F/l. \tag{6.1}$$

Fig. 60

Fig. 61

The physical meaning of this quantity is as follows. It is equal to the force of surface tension per unit length of the surface boundary.

The surface tension depends on the nature of media in contact and on temperature. It is expressed in newtons per metre (N/m).

As the temperature increases, the difference between a liquid and its saturated vapour gradually diminishes and at the critical temperature disappears completely. Accordingly, the surface tension for the liquid-saturated vapour interface decreases with increasing temperature and becomes zero at the critical temperature.

According to formula (6.1), the magnitude of the force of surface tension acting on the boundary of the surface layer of length l is given by

$$F = \sigma l. \tag{6.2}$$

The force of surface tension is directed along the tangent to the surface at right angles to the interface (i.e. normally to the wire AB in the case under consideration, see Fig. 60 and 61). This force is proportional to the length, the proportionality factor being the surface tension.

6.3. Capillary Phenomena

WETTING. This phenomenon is observed where a liquid surface is in contact with solids. *Wetting is a phenomenon that takes place due to the interaction between the molecules of a liquid and a solid and is manifested in bending the liquid surface at the surface of the solid.*

We can neglect the attraction between molecules of a liquid and gas molecules, but the interaction between molecules of liquids and solids cannot be ignored. The shape of the liquid surface in contact with a solid depends on whether the attraction between molecules of the solid and the liquid is stronger than the attraction of the molecules of the liquid itself.

If the attraction between the molecules of a solid and a liquid is stronger than that of liquid molecules, the liquid wets the solid, and the shape of the surface (Fig. 62) is such that the angle θ between the plane tangential to the surface of the liquid and the vessel wall is acute (the liquid sticks to the wall). Otherwise, the liquid does not wet the surface of the solid, and the angle θ (Fig. 63) is obtuse (the liquid is "repelled" from the wall).

If wetting is complete, the angle θ = 0°, while at zero wetting θ = 180°. Wetting plays a significant role in engineering and in everyday life. Good
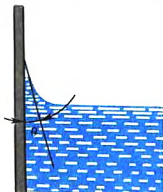

Fig. 62   Fig. 63
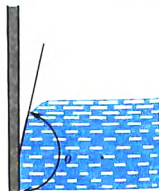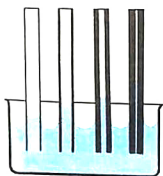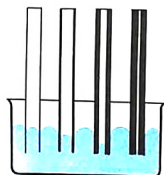
Fig. 64

Fig. 65

Fig. 66

wetting is required in dying, washing, processing photographic materials, painting and varnishing, and so on. Glueing of wooden, leather, rubber and other materials is also an example where wetting is employed. Wetting is also used in soldering. The molten solder will flow uniformly over the surface of metallic objects and stick to them only if these surfaces are cleansed of fat, dust and oxides. Tin-based solders are quite effective for soldering objects made of copper, brass, etc. This solder, however, does not wet aluminium. A special solder is employed for aluminium and silicon.

CAPILLARY PHENOMENA. *Capillary phenomena include rising or lowering of a liquid in narrow tubes called capillaries* [1] as compared to the level of the liquid in wide tubes.

A wetting liquid (like water in a glass tube) rises in a capillary. The smaller the radius of the tube, the larger the height to which the liquid rises (Fig. 64). If we look at the meniscus of such a liquid in a capillary through a magnifying glass, it resembles a stretched rubber film fixed to the walls of the tube and depressed at the middle (Fig. 65).

In the case of complete wetting or nonwetting, the surface of a liquid in a capillary can be regarded as a hemisphere whose radius is equal to the radius r of the channel of the tube.

A downward force of surface tension (6.2) acts along the circular boundary of the surface layer. According to Newton's third law, a force of the same magnitude is exerted by the walls on the liquid in the upward direction. This force makes the liquid rise. Since the length of the circle is $l = 2\pi r$, this force is given by

$$F = \sigma 2\pi r. \tag{6.3}$$

A liquid stops rising in a capillary when the force (6.3) is balanced by the force $m\vec{g}$ of gravity acting on the liquid column:

$$\vec{F} = -m\vec{g}. \tag{6.4}$$

We denote by h the height to which the liquid rises in the capillary. Then the volume of the liquid column (cylinder) is $V = \pi r^2 h$. The magnitude of the force of gravity is

---

[1] The term capillary takes its origin from the Latin word *capillus* meaning "a hair".
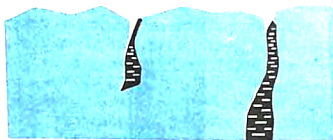
Fig. 67

$$mg = \rho g V = \rho g h \pi r^2, \tag{6.5}$$

where ρ is the density of the liquid.

Substituting expressions (6.3) and (6.5) into Eq. (6.4), we obtain

$$2\sigma\pi r = \rho g h \pi r^2.$$

Hence the height to which a wetting liquid rises in a capillary is

$$h = \frac{2\sigma}{\rho g r}. \tag{6.6}$$

The level of a liquid which does not wet a capillary (like mercury in a glass tube) is below the level of the liquid in a wide vessel (Fig. 66). The depth h to which the liquid is lowered is also determined by formula (6.6).

Bodies pierced by a large number of thin channels (capillaries) actively absorb water and other liquids. It is only necessary that the liquids wet the surface of the body. Just for this reason, a towel absorbs water when we dry hands with it. Kerosene continuously rises in the capillaries of the wick of a kerosene stove or lamp and is burnt at the top.

The wick method is sometimes used for supplying a lubricant (oil) to the parts of machines.

Ordinary bricks are porous bodies that are good absorbers of moisture. For this reason, the foundations of brick buildings should be protected from moisture. This is done by covering them with hot bitumen or tar paper.

Water in the soil rises in its capillaries to the surface where it evaporates vigorously (Fig. 67).

?
1. Give example of operating forces of surface tension, which have not been mentioned in the text.
2. Compare and contrast the forces of surface tension and elastic forces.
3. What is surface tension?
4. What does the surface tension depend on?
5. What is the force of surface tension?
6. How is the height to which a liquid rises in a capillary determined?

Problems with Solutions

The problems to this chapter differ from ordinary problems on hydrostatics only in that here we also take into account one more force, viz. the force of surface tension (6.2).

96

In many problems, formula (6.6) is used, which defines the height to which a liquid rises in a capillary.

1. A cube of mass $m = 0.02$ kg floats on the surface of water wetting it. The edge of the cube is $a = 0.03$ m. At what distance x from the surface of water is the lower face of the cube?

Solution. The upward buoyancy force $a^2 x \rho \hat{g}$ is balanced by the force of gravity $m\hat{g}$ acting on the cube and the force of surface tension. According to (6.2), the force of surface tension is equal to $4a\sigma$ and is directed downwards. Consequently, for the projections of the forces on the upwards-directed x-axis, we have

$$a^2 x \rho g - mg - 4a\sigma = 0.$$

Hence

$$x = \frac{mg + 4a\sigma}{a^2 \rho g} \simeq 0.023 \text{ m}.$$

The correction due to the force of surface tension is of the order of 1 mm.

2. A vertical capillary tube with an inner radius $r = 10^{-4}$ m is immersed in mercury. Assuming that mercury does not wet the material of the tube at all, determine the pressure of mercury in the capillary directly under the convex (hemispherical) surface of the liquid. The atmospheric pressure should be neglected.

Solution. It follows from the equilibrium condition for the liquid column in a tube that the pressure below the liquid surface is equal to the pressure at the same level outside the tube. This pressure is equal to $p = \rho g h$, where h is the depth of the level of mercury in the tube. Using the expression (6.6) for h, we obtain

$$p = \frac{2\sigma}{r} = \frac{9.4 \times 10^{-1}}{10^{-4}} \text{ Pa} \simeq 9400 \text{ Pa}.$$

Exercise 6

1. A teaspoon of olive oil is poured into an aqueous solution of alcohol (Plateau's experiment). What is the shape of oil if it has the same density as the solution?
2. Verify experimentally whether the surface tension of water increases or decreases as a result of dissolving sugar or soap in it.
3. The surface tension of a solution of grease in petrol is larger than for pure petrol. Should the centre of a grease stain or its boundaries be wetted by petrol for stain removal?
4. A wire whose diameter is 0.1 mm is suspended vertically over the pan of an analytic balance and is partially immersed in a vessel containing water. The surface tension of water is 7.3×10^{-2} N/m. Find the additional force acting on the balance due to wetting the wire by water.
5. Why does the harrowing of soil help to preserve moisture in it?
6. Alcohol rises in a capillary tube to a height of 55 mm, while water rises to 146 mm in the same tube. Determine the density of alcohol.

Main Points of Chapter 6

The force of surface tension is the force acting along the surface of a liquid normally to the boundary of this surface and tending to reduce it to the minimum possible value. The magnitude of the force of surface tension acting on a surface layer boundary of length l is

$$F = \sigma l,$$

where σ is the surface tension. This factor depends on the nature of the media in contact and on the temperature.

Owing to the forces of surface tension, liquids wetting the surfaces of thin tubes (capillaries) rise, while liquids which do not wet the surface of such tubes go down. The height by which liquids rise (or fall) is given by

$$h = 2\sigma/\rho g r,$$

where ρ is the density of a liquid, g is the free fall acceleration and r is the radius of the capillary.

7 SOLIDS

7.1. Crystalline Bodies

SOLID BODIES not only retain their volume like liquids, but also preserve their shape. Most solids are in the CRYSTALLINE STATE.

Crystals are solids whose atoms or molecules occupy definite ordered positions in space (see Sec. 1.5). As a consequence, crystals have a regular shape. For example, a grain of common salt has flat faces at right angles to one another (Fig. 68). This can be noticed by observing common salt through a magnifying glass. Snow flakes have surprisingly regular geometrical shapes. The shape of a snow flake reflects the geometrical regularity of the internal structure of ice crystals (Fig. 69).

ANISOTROPY OF CRYSTALS. However, the regular shape is not the only and not even the main consequence of the ordered structure of crystals. The main feature of crystals is the dependence of their physical properties on a selected direction in a crystal. The most striking is the fact that the mechanical strength of a crystal is different in different directions. A piece of mica can be easily stratified into thin laminas (Fig. 70). It is much more difficult to tear a mica sample in a direction perpendicular to the laminas. Similarly, a graphite crystal easily delaminates in one direction. When we write with a pencil, this foliation occurs continuously, and thin layers of graphite remain on the paper. This is so because the crystal lattice of graphite has a laminated structure. The layers are formed by a number of parallel plane grids consisting of carbon atoms (Fig. 71). The atoms are arranged at
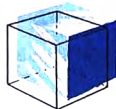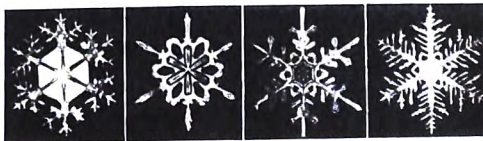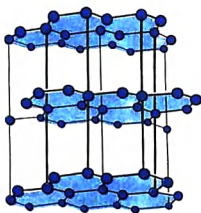

Fig. 68   Fig. 69
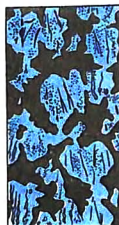
Fig. 70

Fig. 71

Fig. 72

the vertices of regular hexagons. The separation between the layers is comparatively large: it is twice the side of the hexagon. For this reason, the bonds between the layers are weaker than the bonds within them.

Many crystals have different values of thermal conductivity in different directions. The optical properties of crystals also depend on direction. For example, quartz crystals refract light depending on the direction of incident rays.

The dependence of physical properties on the directions in a crystal is known as anisotropy [1]. *All crystals are anisotropic.*

SINGLE CRYSTALS AND POLYCRYSTALS. Metals have a crystalline structure. They are mainly used at present for manufacturing instruments. machine tools and parts of various mechanisms.

If we take a piece of metal, its crystalline structure is not manifested at first sight either in its appearance or in its physical properties. Under normal conditions, metals do not exhibit anisotropy.

As a matter of fact, a metal normally consists of a huge number of small crystals grown together. In a fresh fracture of a metal (Fig. 72), they can be seen through a microscope or even through a magnifying glass. The properties of each small crystal are different in different directions, but the crystals are oriented at random relative to one another. As a result, in a volume considerably exceeding the volume of an individual crystal all directions are equivalent, and the properties of a metal are uniform in all directions.

A solid consisting of a large number of small crystals is called a POLYCRYSTALLINE solid. Isolated crystals are known as SINGLE CRYSTALS (or MONOCRYSTALS).

With considerable precautions, a metallic single crystal of a large size can be grown. Under normal conditions, a polycrystalline solid is formed as a result of merging of many crystals whose growth is terminated when they come into contact with one another.

Polycrystals include not only metals. A piece of sugar, for example, also has a polycrystalline structure.

----

[1] From the Greek words *anisos* meaning "unequal" and *tropos* meaning "direction".
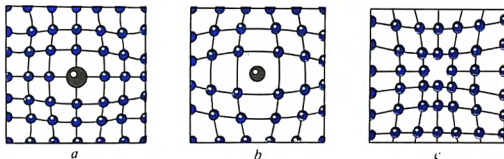
Fig. 73

DEFECTS IN CRYSTALS. In all real crystals, distortions of strictly periodic structure in atomic arrangement are observed. These distortions are termed crystal defects. Defects in crystals appear in the process of their growth as a result of thermal motion of molecules, mechanical effects, irradiation by particle beams, and so on.

The simplest type of crystal defects are the POINT DEFECTS. Such defects appear as a result of substitution of an atom in a crystal lattice by an impurity atom (Fig. 73a), a foreign atom intrusion between lattice sites (Fig. 73b) or the formation of a vacancy, viz. a missing atom in a lattice site (Fig. 73c).

LINEAR DEFECTS, viz. distortions in the lattice structure concentrated along straight lines, produce a much stronger effect on the mechanical and other properties of materials. These defects are known as DISLOCATIONS (displacements).

Atomic planes (layers) often terminate in the bulk of a crystal. As a result, an additional half-plane is formed (Fig. 74). Such a defect is called EDGE dislocation.

SCREW dislocations represent more complex defects. They can be visualized as a result of "cutting" a lattice along a half-plane (Fig. 75) followed by a displacement of the parts of the lattice, situated on the opposite sides of the cut, by an interatomic distance. A crystal with a screw dislocation is formed not by parallel atomic layers but rather is represented by one plane twisted in the form of a spiral staircase. Such a crystal grows by attaching atoms to the edge of the last run. Therefore, the crystal layer grows by
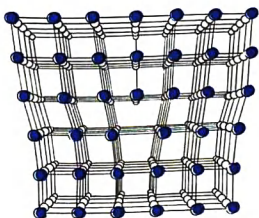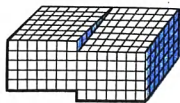

Fig. 74

Fig. 75

Fig. 76

101

continuously winding on itself, and the crystal acquires the shape of a screw (Fig. 76).

The number of dislocations in a crystal may vary over a wide range from 10^2-10^3 cm^{-2} to 10^{11}-10^{12} cm^{-2}.

The growth of defectless crystal is a very complicated problem which has been solved only for a few substances.

The knowledge of conditions for formation of defects in crystals and methods of their elimination is very important for the application of crystalline bodies in technology (see Sec. 7.5).

7.2. Amorphous Bodies

Along with crystalline solids, amorphous solids [1] are also encountered. Unlike crystals, amorphous bodies have no strict order in the arrangement of atoms. Only closest atoms, viz. neighbours, are arranged in a certain order. However, a strict repetition of the same element of the structure, which is typical of crystals, is not observed in amorphous bodies.

The same substance may be either in the crystalline or in the amorphous state. For example, quartz (SiO_2) is found both in the crystalline form and in the amorphous state (silica). The crystalline form of quartz can be schematically represented as a lattice of regular hexagons (Fig. 77a). The amorphous form of quartz is also represented by a lattice but of irregular shape. In addition to hexagons, it contains pentagons and heptagons (Fig. 77b).

PROPERTIES OF AMORPHOUS BODIES. All amorphous bodies are isotropic [2]. Their physical properties are identical in all directions. Amorphous substances include glass, many plastics, resin, colophony, sugar candy and others.

Under the influence of external agencies, amorphous bodies exhibit simultaneously elastic properties like solids and fluidity like liquids. Under short-term effects (impacts), they behave like solids and break into small pieces under a strong blow. However, under a long-term action on them, they flow. For example, a piece of resin gradually spreads over a solid surface. Atoms and molecules of amorphous bodies, like liquid molecules, have a definite time of "settled life", viz. the time of vibrations about an equilibrium position. In contrast to liquids, this time is rather long for amorphous bodies. In this respect, they are close to crystalline bodies since the atoms rarely jump from one equilibrium position to another.

The properties of amorphous bodies at low temperatures resemble the properties of solids. They have nearly zero fluidity. But as the temperature increases, amorphous bodies become soft, and their properties become closer and closer to the properties of liquids. This is due to the fact that the atoms

[1] From the Greek *a* meaning "without" and *morphe* meaning "form".

[2] From the Greek *isos* meaning "equal" and *tropos* meaning "direction".
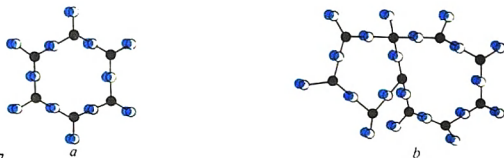
Fig. 77

jump more frequently from one equilibrium position to another with increasing temperature. There is no definite fusion temperature for amorphous bodies in contrast to crystalline bodies.

SOLID STATE PHYSICS. Solids have been and will be used by man. But unlike old times when the solid state physics lagged behind the rapidly developing technology based on direct experience, the situation is quite different now. Theoretical studies now lead to the production of solids with quite extraordinary properties, which could not be obtained by the "trial-and-error" method. A brilliant example of such materials are transistors which will be described below. In this case, the knowledge of the structure of solids resulted in a revolution in radio engineering.

The creation of materials with preset mechanical, magnetic or electrical properties is one of the main trends in the modern solid state physics. About half the number of physicists all over the world are engaged in this branch of physics.

? 1. What is the difference between crystalline and amorphous bodies?
2. What is anisotropy?
3. Give examples of single crystals, polycrystalline and amorphous bodies.

7.3. Deformation. Types of Deformation in Solids

DEFORMATION OF A SOLID BODY. *Deformation is a change in the shape or volume of a body.*

Deformation takes place when different parts of a body have different displacements. For example, if we stretch a rubber cord at its ends, the parts of the cord are displaced relative to one another, and the cord becomes deformed (it becomes longer and thinner).

A deformation that completely vanishes after the action of external forces is eliminated is known as ELASTIC. An example of elastic deformations is that of a string which restores its original shape after the load suspended to it has been removed.

A deformation that does not disappear after the external forces have been withdrawn is known as PLASTIC deformation. This type of deformation can be observed in wax, plasticine, clay or lead even under small (but prolonged) effects.

All types of deformations in solids can be reduced to two kinds, viz. an extension (compression) and shear.
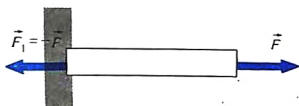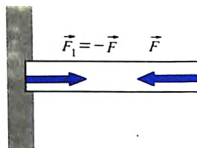
103

Fig. 78   Fig. 79

EXTENSION (COMPRESSION). If we apply a force $\vec{F}$ to a homogeneous rod fixed at one end along its axis and away from it (Fig. 78), the rod experiences deformation due to extension. Extension is characterized by the ABSOLUTE ELONGATION $\Delta l = l - l_0$ and the RELATIVE ELONGATION

$$\varepsilon = \Delta l / l_0,$$

where l_0 is the initial length and l is the final length of the rod.

Cables, ropes, chains in hoisting equipment, and car couplers undergo extension.

For small extension ($\Delta l \ll l_0$), the deformations are mostly elastic.

If a force $\vec{F}$ acts along the axis of a rod fixed at one end (Fig. 79) towards the fixed end, the rod undergoes compression. In this case, the relative deformation (strain) ε is negative: $\varepsilon < 0$.

Pillars, columns, walls and foundations of buildings undergo this type of deformation.

In extension or compression, the cross-sectional area of a body changes. This can be observed by stretching a rubber tube with a metallic ring which has been put on it beforehand. When the extention is considerable, the ring falls down. On the contrary, the cross-sectional area of a body increases upon compression. For most solids, however, these effects are small.

SHEAR. Let us take a rubber block and draw vertical and horizontal lines on its surface. We fix this block on the table (Fig. 80a), attach a strip to its upper face, and apply a horizontal force to it (Fig. 80b). The layers ab, cd, etc. of the block are displaced but remain parallel to one another, while the
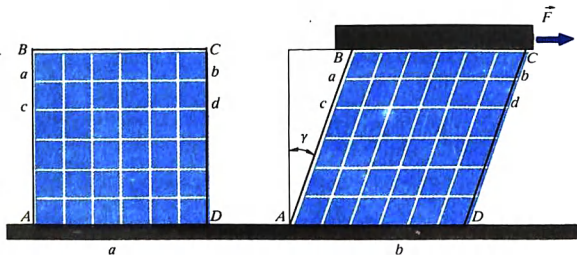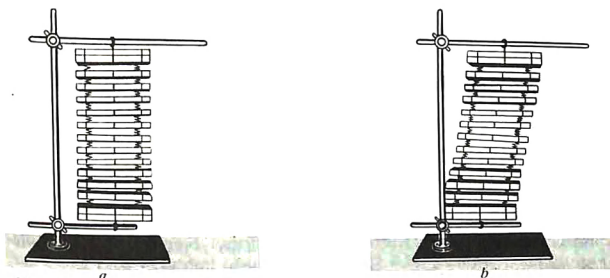

Fig. 80

104

Fig. 81

vertical faces remain plane but are inclined by an angle γ. This type of deformation in which the layers of a body are displaced relative to one another is known as SHEAR.

If the force $\vec{F}$ is doubled, the angle γ also increases twofold. Experiments show that *in elastic deformations, the angle of shear γ is proportional to the magnitude F of the applied force.*

Shear can be visually demonstrated by using a model of solid, which is in the form of parallel plates interconnected by springs (Fig. 81a). A horizontal force shifts the plates relative to one another without changing the volume of the body (Fig. 81b). In the shear of real bodies, their volume also remains unchanged.

Shear is observed in beams at contacts with the supports, in rivets (Fig. 82) and bolts connecting parts of machines, etc. Shear at large angles may lead to a breakdown of a body, viz. SHEARING. Shearing takes place in the operation of scissors, chisel, chipper, and the teeth of a saw.

BENDING. This type of deformation is observed in a rod supported at its ends and loaded at the middle, or a rod fixed at one end and loaded at the other (Fig. 83).

In bending, one face (convex) undergoes extension and the other (concave) undergoes compression. Inside the body being bent, there is a layer that experiences neither extension nor compression. This layer is known as the NEUTRAL LAYER (Fig. 84).

Thus, bending is a deformation that can be reduced to extensions and compressions which are different in different parts of a body.

In the vicinity of the neutral layer, the body is almost free of deformations. The forces emerging in deformation are also weak in this layer. This means
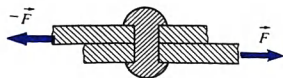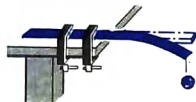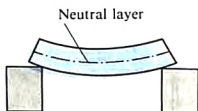

Fig. 82

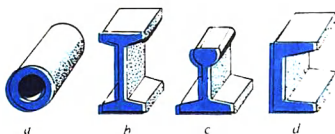Fig. 83

Neutral layer

Fig. 84

Fig. 85

that in the vicinity of the neutral layer, the cross-sectional area of a body undergoing bending can be considerably reduced. In modern engineering and construction of buildings, tubes (Fig. 85a), flange beams (Fig. 85b), rails (Fig. 85c) and channels (Fig. 85d) are widely used instead of rods and solid bars. This makes the structures lighter and saves the material.

TORSION. If a rod fixed at one end is acted upon by two parallel and opposite forces $\vec{F}_1$ and $\vec{F}_2$ (Fig. 86) lying in a plane normal to the rod axis, a deformation known as TORSION takes place. In torsion, as in shear. the layers of a body remain parallel but are turned by a certain angle relative to one another. Torsion is a deformation equivalent to nonuniform shear.

This deformation emerges, for example. while tightening a nut (Fig. 87). Lathe shafts and drills also undergo torsion.

7.4. Mechanical Properties of Solids. Stress-Strain Diagram

In order to construct sturdy houses, bridges, lathes and various machines, it is important to know the mechanical properties of the materials used for this purpose, viz. concrete. steel, reinforced concrete, plastics, etc. A designer should know beforehand the behaviour of the materials under considerable deformations and the conditions in which the materials suffer fracture. The mechanical properties of various materials are learnt from experiments and tests.

In this section, we shall consider the mechanical properties of solids under tension by way of an example. For this, we shall need one more important concept.

STRESS. In any cross section of a deformed body, elastic forces act to prevent the body from fracture (Fig. 88). The state of the deformed body is characterized by a special quantity called STRESS or, to be more precise,
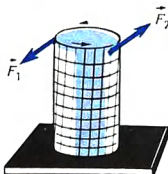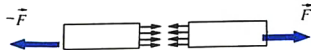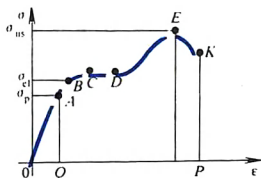

Fig. 86

Fig. 87

Fig. 88

Fig. 89

mechanical stress. Stress is equal to the ratio of the magnitude F of the elastic force to the cross-sectional area S of the body:

$$\sigma = F/S. \tag{7.1}$$

The SI unit of stress is a pascal (1 Pa = 1 N/m²).

STRESS-STRAIN DIAGRAM. The tension of the rod made of a material under investigation is analyzed by stretching the rod in a special device and by measuring the stress emerging in it. Using the results of experiments, the stress σ is plotted as a function of strain ε. This curve is known as the STRESS-STRAIN DIAGRAM (Fig. 89).

HOOKE'S LAW. Experiments show that at small deformations, the stress σ is proportional to the strain ε (region OA in Fig. 89). This dependence is called Hooke's law and written as follows:

$$\sigma = E|\varepsilon|. \tag{7.2}$$

In this formula, the absolute value of strain is taken, since Hooke's law is valid for extension as well as for compression (when $\varepsilon < 0$).

The proportionality factor E in Hooke's law is known as the MODULUS OF ELASTICITY, or YOUNG'S MODULUS. Young's modulus is determined from formula (7.2) by measuring the stress σ and the strain ε for small deformations.

For most of the materials widely used in technology, Young's modulus is determined experimentally. For example, $E = 2.1 \times 10^{11}$ Pa for nickel-chromium steel, while $E = 7 \times 10^{10}$ Pa for aluminium. The larger Young's modulus E, the smaller the deformation of the rod, other conditions being equal (i.e. for the same F, S and l_0). Young's modulus characterizes the resistance of a material to elastic tension (compression).

Hooke's law (7.2) can be easily reduced to the form familiar from the course of "Senior Physics 1". Indeed, substituting $\sigma = F/S$ and $\varepsilon = |\Delta l|/l_0$ into (7.2), we obtain

$$F/S = E|\Delta l|/l_0.$$

Hence

$$F = \frac{SE}{l_0}|\Delta l|. \tag{7.3}$$

We denote SE/l_0 by k. This gives

$$F = k|\Delta l|. \tag{7.4}$$

107

Thus, the rigidity k of the rod is directly proportional to the product of Young's modulus and the cross-sectional area of the rod and is inversely proportional to its length.

LIMITS OF PROPORTIONALITY AND ELASTICITY. As was mentioned above, Hooke's law is valid for small deformations, and hence for stresses below a certain limit. The maximum stress σ_p for which Hooke's law is still observed is called the proportional limit (or limit of proportionality).

If the load is increased, the deformation becomes nonlinear: the stress is no longer proportional to the strain. Nevertheless, at small nonlinear deformations, the shape and size of the body are practically restored after the load has been removed (region AB in Fig. 89). The maximum stress at which no noticeable residual deformation appears (the residual strain is below 0.1%) is called the elastic limit σ_{el}. The limit of elasticity exceeds the proportional limit only by a few hundredths of one percent.

ULTIMATE STRENGTH. If the external load is such that the stress in the material exceeds the elastic limit, then although the sample contracts after the removal of the load, it does not resume its original size and remains deformed.

As the load increases, the residual deformation becomes larger and larger. At a certain value of stress corresponding to point C on Fig. 89, the elongation increases without the load being increased. This phenomenon is called the YIELD of the material (region CD). The curve on the diagram is almost horizontal in this region.

As the strain ε increases further, the stress-strain curve begins to rise and attains a maximum at point E. Then the stress rapidly drops, and the sample fractures (point K). Thus, *the fracture occurs after the stress has attained the maximum value σ_{us} known as the ultimate strength* without any increase in the load, (the sample is stretched to fracture). This value depends on the material of the sample and its processing.

7.5. Plasticity and Brittleness

A body made of any material behaves as an elastic body under small deformations. Its size and shape are restored after the load has been removed. On the other hand, almost all bodies may undergo plastic deformations to a certain extent.

Materials have a wide range of mechanical properties. Steel or rubber, for example, exhibit elastic properties at comparatively large stresses and strains. Hooke's law is valid up to $\varepsilon = 1\%$ for steel and at considerably higher values of ε for rubber (of the order of tens percent). For this reason, such materials are called ELASTIC.

PLASTICITY. Wet clay, plasticine and lead have small regions of elastic deformations. *Materials in which insignificant loads cause plastic deformations are called plastic materials.*

The division of materials into elastic and plastic is conditional to a considerable extent. Depending on stresses emerging in it, the same material behaves as an elastic or a plastic material. For example, steel

Fig. 90

exhibits plastic properties under very high stresses. This is widely used in stamping steel articles with the help of presses developing a huge load.

Cold steel or iron can be hardly forged by a hammer. After intensive heating, however, they can be given any shape by forging. Lead which is plastic at room temperature acquires clearly pronounced elastic properties when cooled even from a temperature below -100 C.

BRITTLENESS. This property of solids plays a significant role in practice. *A material is called brittle if it fractures at small deformations.* Glass and porcelain articles are very brittle: they break into pieces when dropped to the floor even from a small height. Cast iron, marble, and amber also are highly brittle. On the contrary, steel, copper, lead, etc. are not brittle at all.

In all brittle materials, stress increases with strain very rapidly, and they break down at quite small deformations. For example, the fracture of cast iron is observed at a strain of 0.45%. For the same $\varepsilon = 0.45\%$, the deformation in steel remains elastic, and fracture occurs at $\varepsilon \simeq 15\%$.

Brittle materials practically do not exhibit plastic properties.

MECHANISM OF PLASTIC DEFORMATIONS. During elastic deformations occurring in crystalline bodies, atoms are displaced relative to one another only slightly.

In plastic deformations, the displacements of atoms and molecules may exceed many times the interatomic and intermolecular distances. However, no distortion of the crystalline structure is observed. Separate layers of a crystal lattice slip relative to one another. Figure 90 shows a small copper crystal after extension. The traces of slips are clearly seen.

It is important that the slip of atomic layers in all crystals begins in separate regions and not over the entire volume of the crystal. It starts at the sites where the crystal lattice has a lower strength for some reason or other. For instance, separate randomly arranged groups of atoms may exist in a crystal.

During plastic deformation in polycrystalline bodies, the slip of layers in the lattice also occurs in separate small crystals. However, the directions of the easiest slip are different in different crystals since they are randomly oriented relative to one another. For this reason, plastic deformations in polycrystalline bodies start at higher stresses than in single crystals. This is observed especially clearly for materials having a fine-grain structure.

?
1. List the main types of deformations in solids.
2. What is mechanical stress?
3. What is the relation between the mechanical stress and strain (relative elongation)?
4. What is the difference between elastic and plastic deformations?
5. Which quantity is known as the ultimate strength?
6. Which materials are called brittle?

Problems with Solutions

In the computational problems to this chapter, the concept of stress (formula (7.1)) and Hooke's law in the forms (7.2) and (7.3) are used.

1. Calculate the relative contraction of a steel rod compressed by a force $F = 3.14 \times 10^5$ N if the rod diameter $D = 2$ cm, and Young's modulus for steel is $E = 2 \times 10^{11}$ Pa.

Solution. According to Hooke's law (7.3), we have

$$|\varepsilon| = \frac{|\Delta l|}{l_0} = \frac{F}{ES},$$

where $S = \pi D^2/4$ is the cross-sectional area of the rod. Consequently, $|\varepsilon| = 4F/\pi D^2 E = 0.005$, or $|\varepsilon| = 0.5\%$.

2. A brass wire has a diameter $D = 0.8$ mm and a length $l = 3.6$ m. Under the action of force $F = 25$ N, the length of the wire increases by $\Delta l = 2$ mm. Determine Young's modulus for brass.

Solution. Young's modulus can be determined from Hooke's law (7.3): $E = Fl_0/(S \cdot |\Delta l|)$. Since $S = \pi D^2/4$, then $E = 4Fl_0/(\pi D^2 \cdot |\Delta l|) \simeq 9 \times 10^{10}$ Pa.

Exercise 7

1. If a body is anisotropic, does this mean that it is necessarily a crystalline body?
2. Would the profession of glass blowing exist if glass were crystalline and not amorphous?
3. A plough is connected to a tractor through a steel coupling bar. The admissible stress for steel of which the bar is made is $\sigma = 20$ GPa. What should be the cross-sectional area of the bar if the resistance offered by soil to the motion of the plough is 1.6×10^5 N?
4. What should be the magnitude of the force applied to a rod along its axis to create in it a stress of 1.5×10^8 Pa? The diameter of the rod is 0.4 cm.
5. What stress emerges at the base of a brick wall of 20-m height if the density of brick is 1800 kg/m^3? Should the strength of the bricks be the same at the base of the wall and in its upper part?
6. What should be the minimum length of a steel wire suspended freely at its end, for it to break down under the action of the force of gravity? The ultimate strength for steel is 3.2×10^8 Pa, and the density of steel is 7800 kg/m^3.
7. A 5-m long wire having a cross-sectional area of 2.5 mm^2 is elongated by 1 mm under the action of a force of 100 N. Determine the stress in the wire and Young's modulus.
8. A reinforced concrete column is compressed by a force F. Determine the fraction of load acting on concrete, assuming that Young's modulus E_c for concrete is 1/10 of Young's modulus E_s for steel, while the cross-sectional area of steel is 1/20 that of concrete.

Solids exist mainly in the crystalline state. Crystals are anisotropic, which means that their physical properties depend on the chosen direction.

Unlike crystals, amorphous bodies do not have a strictly ordered arrangement of atoms. Amorphous bodies resemble solids in their properties at low temperatures and are similar to high-viscosity liquids at high temperatures.

For small deformations, a body resumes its shape and size after the load has been removed. Solid bodies exhibit elastic properties. In elastic deformations, the stress, viz. the ratio of the magnitude of elastic force to the cross-sectional area, obeys Hooke's law:

$$\sigma = E \, | \varepsilon |.$$

where E is Young's modulus characterizing elastic properties of a substance, and $\varepsilon = \Delta l / l_0$ is the relative change in the size of the body and is called strain.

The mechanical properties of solids are clearly represented by the stress-strain diagram (Fig. 89).

Fundamentals of Electrodynamics

What is Electrodynamics?

We start a new part of physics called electrodynamics. The name itself indicates that we shall deal with processes which depend on the motion and interaction of electrically charged particles. This type of interaction is called ELECTROMAGNETIC interaction. The analysis of this interaction will lead us to one of the fundamental concepts in physics, viz. the ELECTROMAGNETIC FIELD.

Electrodynamics is the science that studies the properties and regularities in the behaviour of a special kind of matter – electromagnetic field through which the interaction between electrically charged bodies or particles is realized.

Among the four types of interactions discovered by scientists, viz. gravitational, electromagnetic, strong (nuclear) and weak[1], it is the electromagnetic interactions that are manifested most widely and diversely. In everyday life and in engineering, we frequently encounter various electromagnetic forces. These are elastic forces, friction, the forces developed by muscles of human beings and animals.

Electromagnetic interactions make it possible to see the book you are reading now since light is a form of electromagnetic field. Life cannot be imagined without electromagnetic forces. The flights of cosmonauts showed that living organisms, including man, can exist for a long time in the zero-gravity state when the forces of universal gravitation have no effect on active life of organisms. If, however, the action of electromagnetic forces vanished for a moment, life would cease at once.

Electromagnetic forces play an important role in the interaction of particles in the smallest systems in nature, viz. atomic nuclei, as well as in the interaction of celestial bodies. However, strong and weak interactions determine the processes only on a very small scale[2], and the gravitational

---

[1] Weak interactions are mainly responsible for transformations in elementary particles.

[2] The distances at which strong interactions are observed are of the order of 10^{-12} cm, while weak interactions are manifested at still shorter distances which do not exceed 10^{-17} cm.

interactions are manifested to the greatest extent on the scale of celestial bodies. The structure of atomic shells, binding of atoms into molecules (chemical bonds) and formation of macroscopic bodies are determined exclusively by electromagnetic forces.

Electrodynamics was the culmination of a long chain of planned investigations and accidental discoveries, starting with the discovery that amber rubbed against wool can attract light objects and finishing with the hypothesis put forward by the great English scientist J.C. Maxwell about the generation of magnetic field by a varying electric field. The wide practical application of electromagnetic phenomena became possible only at the end of the 19th century following the advent of electrodynamics. The invention of radio by A.S. Popov was one of the most important applications of the new theory.

It was for the first time in the evolution of electrodynamics that theoretical studies were ahead of technical applications. While the steam engine had been constructed long before the theory of thermal processes was created, the electric motor or radio receiver became possible only after the laws of electrodynamics had been discovered and studied.

Numerous practical applications of electromagnetic phenomena have changed the life of the mankind all over the world. Modern civilization cannot be imagined without the widest application of electric energy.

Our aim is to study the basic laws of electromagnetic interactions and to gain the idea of the principles of production of electric energy and its practical application.

8 ELECTROSTATICS

8.1. Electric Charge and Elementary Particles

We encounter the words "electricity", "electric charge", and "electric current" very often and are accustomed to these concepts. But if you try to answer the question: "What is electric charge?", you see that this is not so simple as it would appear. As a matter of fact, the concept of charge is a primary, basic concept which cannot be reduced to simpler, more elementary concepts on the present level of our knowledge.

Let us try to find out what is meant by saying that a body or particle has an electric charge.

We know that all bodies are composed of tiny particles that cannot be divided into simpler particles (as far as we know at the moment) and are hence called elementary particles. All elementary particles have a mass and thus attract one another in accordance with the law of universal gravitation. The gravitational force decreases with increasing distance between particles at a relatively slow rate: in inverse proportion to the square of the distance between them. Most elementary particles (although not all of them) are able to interact with a force which also decreases in inverse proportion to the squared distance, but this force exceeds the gravitational force enormously. For example, in a hydrogen atom schematically shown in Fig. 91, the electron is attracted to the nucleus (proton) with a force which is 10^{39} times stronger than the gravitational force.

If particles interact with one another through forces that decrease slowly with increasing distance and exceed the forces of universal gravitation many
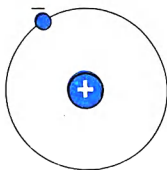

Fig. 91

114

times, they are said to have an ELECTRIC CHARGE and are called *charged particles.* There can be particles without an electric charge, but an *electric charge does not exist without a particle.*

The interaction between charged particles is known as ELECTROMAGNETIC interaction. The electric charge determines the intensity of electromagnetic interaction just as the mass determines the intensity of gravitational interaction.

The electric charge of an elementary particle is not a sort of a special mechanism that can be removed from it, decomposed and reassembled. The presence of an electric charge in the electron and other particles only indicates the existence of force interactions between them. However, we know nothing about a charge unless we know the laws governing these interactions. The knowledge of the laws of interaction should form a part of the concept of charge. These laws are not simple and cannot be explained in a few words. For this reason, we cannot formulate a brief and satisfactory definition of the electric charge.

TWO TYPES OF ELECTRIC CHARGE. All bodies have a mass and hence attract one another. Charged bodies, however, can either attract or repel each other. This important fact described in the course of "Junior Physics" means that there exist particles with unlike charges in nature. Particles having like charges repel each other, while those with unlike charges attract each other.

It is assumed that the charge of elementary particles (protons) contained in all atomic nuclei is positive, while the charge of electrons is negative. There is no internal distinction between positive and negative charges. If the charges of particles changed places, the nature of electromagnetic interactions would remain unchanged.

ELEMENTARY CHARGE. Besides electrons and protons, there exist some other types of charged elementary particles. However, only electrons and protons can exist in free state for an infinitely long time. Other charged particles have lifetimes that are shorter than one millionth of a second. They are born during collisions between fast elementary particles and, having lived for an extremely short time, disintegrate and become some other particles. These particles will be considered later in this course.

The neutron is a particle having zero electric charge. Its mass slightly exceeds the mass of the proton. Neutrons and protons constitute the atomic nuclei.

A large number of experiments (including the Millican and Ioffe experiments described in "Junior Physics") have shown that if an elementary particle has a charge, its value is strictly definite.

There exists a minium charge, which is called the elementary charge and is inherent in all elementary charged particles. The charges of elementary particles differ only in sign. It is impossible, for example, to isolate a fraction of the electron charge.

8.2. Charged Bodies. Electrostatic Charging

Electromagnetic forces play such an important role in nature because electrically charged particles constitute all bodies. The constituents of atoms, viz. nuclei and electrons, have an electric charge.

Electromagnetic forces acting among bodies cannot be perceived directly since bodies are electrically neutral in the normal state. An atom of any substance is also neutral. Positively and negatively charged particles are bound by electric forces and form neutral systems.

A macroscopic body is electrically charged if it contains an excess of elementary particles with one type of charge. The negative charge of a body is due to an excess of electrons over protons, while the positive charge is due to a deficiency of electrons.

In order to impart an electric charge to a macroscopic body, i. e. to CHARGE it electrically, we must isolate a fraction of the negative charge from the positive charge bound to it[1]. This can be done by friction. When we comb dry hair, a small fraction of the most mobile charged particles, viz. electrons, goes over from the hair to the comb and imparts a negative charge to it, while the hair acquires a positive charge.

EQUALITY OF CHARGES IN ELECTROSTATIC CHARGING. We can prove with the help of an experiment that in electric charging by friction the two bodies acquire equal and opposite charges. Let us take an electrometer[2] and fix a hollow metallic sphere to its rod. We shall also need ebonite and plexiglass plates with long handles. The plates are electrically charged by rubbing them against each other. Let us introduce one of the plates into the sphere without touching its walls. If the charge of the plate is positive, a fraction of electrons from the pointer and the rod are attracted to the plate and are collected on the inner surface of the sphere. The pointer and the rod thus become positively charged and repel each other (Fig. 92a). If we introduce another plate into the sphere instead of the first one, the electrons of the sphere and the rod will be repelled from the plate, and the pointer will acquire an excess negative charge. The pointer will be deflected by the same angle as in the first experiment. If we insert the two plates in the sphere simultaneously, the pointer will not deviate at all (Fig. 92b). This proves the fact that the charges on the plates are unlike and have the same magnitude.

MECHANISM OF ELECTROSTATIC CHARGING. In this process, it is important that the bodies are in close contact. Electric forces confine the electrons within a body, but these forces are different in different substances. When the bodies are in close contact, a small fraction of electrons of the

---

[1] Here and below, we shall speak of "charges" and "displacement of charges" for the sake of brevity. Actually, we mean "charged bodies" and "displacement of charged bodies (or particles)", since no charge can exist without a particle.

[2] The construction of an electrometer will be described in detail in Sec. 8.18. The operating principle of an electrometer is the same as that of an electroscope described in the course of "Junior Physics".
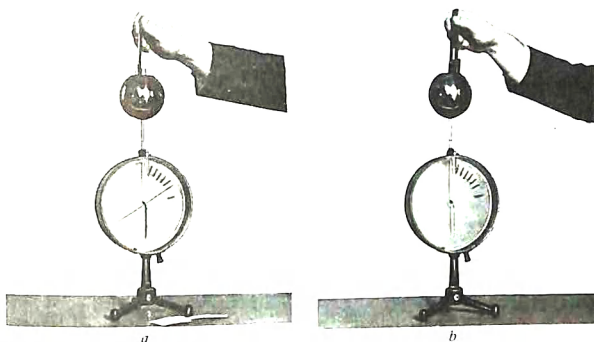
Fig. 92

substance in which the electrons are more weakly bound go over to the
other body. The displacements of electrons in this case do not exceed the
interatomic distances (10^{-8} cm). If we now move the bodies apart, they turn
out to be charged.

Since the surfaces of bodies cannot be perfectly smooth, the contact
between the bodies required for transitions of electrons is limited only to
small regions of the surfaces (Fig. 93). By rubbing one body against the
other, we increase the number of regions where close contact takes place,
and thus the total number of charged particles passing from one body to the
other becomes larger.

ELECTROSTATIC CHARGING AND ITS APPLICATION IN
ENGINEERING. Considerable electrostatic charging occurs when synthetic
cloths are in friction. Taking off a nylon shirt in dry air, you may hear
a typical crackling sound. Tiny sparks run between charged regions of
rubbing surfaces. Such phenomena should be taken into account in industrial
processes. For example, threads in textile factories are electrostatically cha-
rged by friction, are attracted to spindles and rollers and break. Yarn attracts
dust particles and clogs up. Special measures should be taken to prevent
electrostatic charging of threads.

Electrostatic charging of bodies in close contact is used in electrical copying
machines (like Xerox).

These machines operate as follows. A black resinous powder is mixed with very
tiny glass balls. In this process, the balls acquire a positive charge, while the resinous
particles are charged negatively. Due to attraction, they cover the surfaces of the balls
by thin layers.

The text or drawing to be copied is projected onto a thin selenium plate whose
surface is charged positively. The plate rests on a negatively charged metallic surface.
Under the action of light the plate loses its charge, and a positive charge remains only
on the regions corresponding to the latent image sites. Then the plate is covered by
a thin layer of balls. Due to the attraction of unlike charges, resinous powder is

117

attracted to the positively charged regions of the plate. Then the balls are shaken off, and a replica of the image can be obtained on a sheet of paper tightly pressed against the plate. The replica is fixed by heating.

8.3. The Law of Charge Conservation

Experiments with electrostatic charging of plates prove that in charging by friction the charges are redistributed between the bodies that were initially neutral. A small fraction of electrons goes over from one body to the other. No new charged particles appear, nor do the particles that existed before vanish.

In electrostatic charging, the LAW OF CHARGE CONSERVATION is satisfied. This law is valid for a system which does not receive charges from without and does not lose charged particles (such a system is called closed). In a closed system, the algebraic sum of the charges of all the particles remains unchanged. If we denote the charges of the particles by q_1, q_2, ..., we can write

$$q_1 + q_2 + q_3 + ... q_N = \text{const.} \qquad (8.1)$$

The law of charge conservation has a deep meaning. If the number of elementary charged particles does not change, then this law is obviously satisfied. But elementary particles can be transformed into one another, be born or annihilate, giving birth to other particles[1]. In all cases, however, charged particles are generated only in pairs with equal and opposite charges. Charged particles being transformed into neutral particles are also annihilated only in pairs. In all cases, the sum of the charges in an isolated (closed) system remains unchanged.

The validity of the law of charge conservation is confirmed by observing a very large number of transformations that elementary particles undergo. This law expresses one of the fundamental properties of electric charge.

The reason behind the charge conservation is unknown so far.

?
1. Which type of interaction is called electromagnetic interaction?
2. Give the definition of elementary charge.
3. Give examples of phenomena observed by you in everyday life, which are caused by electrostatic charging of bodies.
4. An ebonite rod has acquired a negative charge as a result of

---

[1] This material will be considered in greater detail in the next volume of this course.

electrostatic charging. Has the mass of the rod remained unchanged?

5. Formulate the law of electric charge conservation.
6. Give examples of phenomena in which charge conservation is observed.

8.4. Coulomb's Law – the Basic Law of Electrostatics

Let us carry out a quantitative analysis of electromagnetic interactions. We shall first consider a simple case when electrically charged bodies are at rest. *The part of electrodynamics devoted to static electric charges is known as electrostatics.* The basic law of electrostatics deals with the interaction of two static pointlike charged bodies or particles. This law was established experimentally by the French physicist Ch. Coulomb in 1785 and is called after him.[1]

Pointlike charged bodies do not exist. If, however, the separation of bodies is much larger than their sizes then neither shape nor dimensions of charged bodies affect their interaction noticeably. In this case, the bodies can be treated as point particles. It should be recalled that the law of universal gravitation was formulated for point particles as well.[2]
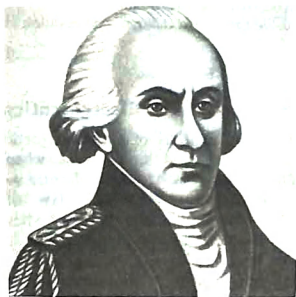
The force of interaction of charged bodies depends on the properties of the medium in which the bodies are placed. We shall assume for the time being that the interaction takes place in vacuum. Experiments show, however, that air affects the force of interaction of charged bodies only slightly: it turns out that this force is practically the same as in vacuum.

COULOMB'S EXPERIMENTS. The discovery of the law of interaction of electric charges was facilitated by the fact that these forces are large. The experiments did not require sensitive instruments like those employed for the verification of the law of universal gravitation in terrestrial conditions. The interaction between fixed charged bodies was studied with the help of a torsion balance. This device consists of a glass rod suspended by a thin elastic wire (Fig. 94). A small metallic ball *a* is attached to one end of the rod, and a counterweight *c* is on the other end. Another metallic ball *b* is fixed at the cover of the balance.

If like charges are imparted to the balls, they start to repel each other. To keep the balls at a fixed distance, the elastic wire must be twisted through a certain angle. The force of interaction of the balls is determined by the angle of twist.

---

[1] In fact, this law was established for the first time by the English scientist Cavendish, but he did not publish his works on electricity. For about a hundred years, the manuscripts were stored in the archives of the Cavendishes until they were transferred to J. Maxwell for publication. This was many years after Coulomb had established the law of interaction of charges.

[2] Knowing the law of interaction of pointlike charged bodies, we can calculate the force of interaction between any charged bodies. For this purpose the bodies must be mentally divided into small elements such that each of them can be regarded as pointlike. By composing the forces of interaction of these elements with one another we obtain the resultant force of interaction between the bodies.

Charles Augustin Coulomb (1736–1806) was a French scientist recognized for his research works in electricity and magnetism as well as investigations of friction. Besides investigating the interaction of charged bodies, Coulomb also studied interaction between the poles of long magnets.

The torsion balance made it possible to study the dependence of the force of interaction between the charged balls on the magnitude of the charges and on their separation. The measurement of force and distance presented no difficulty at Coulomb's time. The only difficulty was associated with the charge for which no units of measurement existed. Coulomb found a simple way of changing the charge of one of the balls two, four, etc. times by connecting it with a similar neutral ball. The charge is distributed in equal parts between the balls, thus reducing the charge under investigation in the known proportion. The magnitude of force corresponding to the new charge was determined experimentally.

COULOMB'S LAW. Coulomb's experiments led to the establishment of a law strikingly resembling the law of universal gravitation. The force of interaction between two fixed point charges in vacuum is directly proportional to the product of the magnitudes of their charges and inversely proportional to the square of the distance between them.[1] This force is known as COULOMB'S FORCE.

If we denote the absolute values of the charges by $|q_1|$ and $|q_2|$, and the distance between them by r, we can write Coulomb's law as follows:

$$F = k\frac{|q_1||q_2|}{r^2}. \tag{8.2}$$

where k is the proportionality factor numerically equal to the force of interaction between unit charges separated by a unit length. The value of k depends on the choice of the system of units.

The law of universal gravitation has a form similar to (8.2) if we take masses instead of charges, the role of k being played by the gravitational constant.

---

[1] Here and below, we shall speak about force instead of the magnitude of force for the sake of brevity.
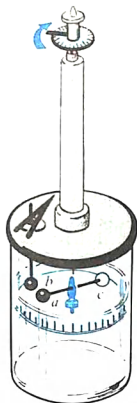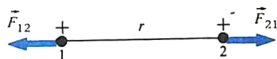
120

Fig. 94

Fig. 95

We have not said anything about the direction of the interaction forces between the charges. It can be easily shown that two charged balls suspended by threads either attract or repel each other. Hence it follows that the forces of interaction of two fixed point charges are directed along the line connecting them (Fig. 95). Forces of this type are called CENTRAL forces. According to Newton's third law, $\vec{F}_{12} = -\vec{F}_{21}$.

The discovery of Coulomb's law was the first step in the study of properties of electric charge. The presence of an electric charge in bodies or elementary particles means that they interact with one another in accordance with Coulomb's law.

8.5. Unit of Electric Charge

The unit of electric charge, like units of other physical quantities, can be chosen arbitrarily. We must only be sure that our choice is expedient.

It is impossible to produce a standard unit of electric charge similar to the standard unit of length (metre) because of the inevitable leakage of charge. It would be natural to take the electron charge as a unit (as is done now in atomic physics). However, the electron charge is too small, and it is often inconvenient to use it as a unit of charge.

COULOMB AS A UNIT OF CHARGE. The unit of charge in the International System of Units (SI) is not a basic unit but a derived one, and no standard is introduced for it. For measuring electric quantities, a base unit, viz. AMPERE, is introduced in SI along with metre, second and kilogram. The standard value of an ampere is established with the help of magnetic interaction of currents. This was mentioned in the course of "Junior Physics".

121

The SI unit of charge, viz. a COULOMB, is defined in terms of the unit of current. One coulomb (C) is the charge passing through the cross section of a conductor carrying a current of one ampere per second.

The coefficient k in Coulomb's law written in terms of SI units has the dimensions of $N \cdot m^2/C^2$, since according to (8.2),

$$k = \frac{Fr^2}{|q_1||q_2|},$$ (8.3)

where the force is expressed in newtons, the distance in metres and the charge in coulombs. The numerical value of this coefficient can be determined experimentally. For this it is sufficient to measure the force of interaction between two known charges at a given distance and substitute the values of F, r, $|q_1|$ and $|q_2|$ into (8.3). The experimentally determined value of k is

$$k = 9 \times 10^9 \ N \cdot m^2/C^2.$$ (8.4)

The charge of one coulomb is very large. The force of interaction between two point charges of one coulomb each separated by one kilometre is slightly less than the force with which the Earth attracts a body whose mass is one tonne. Therefore, it is impossible to impart a charge of 1 C to a small body (whose size is of the order of a few metres). Charged particles repelling one another cannot be confined to a body. There are no other forces in nature that would be capable of balancing the Coulomb repulsion. However, it is not difficult at all to move a charge of 1 C in a conductor which is neutral as a whole. Even in an ordinary electric lamp of 100 W power a current slightly less than 1 A is established at a voltage of 127 V. In this case, a charge almost equal to 1 C passes through the cross section of the conductor.

The minium charge existing in nature is the charge of elementary particles. In SI units, the absolute value of this charge is

$$e = 1.6 \times 10^{-19} \ C.$$ (8.5)

ELECTRIC CONSTANT. The coefficient k in SI is usually written in the form

$$k = \frac{1}{4\pi\varepsilon_0}.$$ (8.6)

The quantity ε_0 is known as the ELECTRIC CONSTANT. Its value is

$$\varepsilon_0 = \frac{1}{4\pi k} = 8.85 \times 10^{-12} \ C^2/(N \cdot m^2).$$ (8.7)

Henceforce, for the sake of simplicity we shall retain in formulas the coefficient k without resorting to its expression (8.6) almost everywhere.

?
1. Compare and contrast the law of universal gravitation and Coulomb's law?
2. How is the unit of charge defined in SI?

3. Formulate Coulomb's law.
4. What is the value of the coefficient k in Coulomb's law?
5. What is the magnitude of the electron charge?

Problems with Solutions

Problems involving Coulomb's law are solved by the same methods as those used for solving problems in statics in the first volume of "Senior Physics". You must only bear in mind that the direction of Coulomb's force depends on the signs of interacting charges. Besides, in some problems use is made of the law of charge conservation and the fact that the minimum portion of electric charge, viz. the absolute value of the charge of elementary particle, is equal to 1.6×10^{-19} C.

1. What is the number of electrons contained in a drop of water of mass $m = 0.03$ g? The mass of a water molecule is $m_0 = 3 \times 10^{-23}$ g.

Solution. The water molecule H_2O contains ten electrons. The drop of water contains $m/m_0 = 10^{21}$ molecules, and hence 10^{22} electrons.

2. Two identical balls are suspended by threads of length $l = 2.0$ m at the same point. When equal charges $q = 2.0 \times 10^{-8}$ C are imparted to the balls, they move apart by a distance $r = 16$ cm. Determine the tension in the threads.

Solution. Each ball experiences the action of three forces: the force of gravity $m\vec{g}$, the elastic force $\vec{F}_{el}$ of the thread, and the Coulomb force $\vec{F}$ (Fig. 96).

The balls are at rest; consequently, the sum of projections of the forces onto the x- and y-axes is zero. For the sum of the x-projections of the forces, this condition has the form: $F - F_{el} \sin \alpha + mg \cos 90° = 0$. Since $\sin \alpha = r/(2l)$ and $F = kq^2/r^2$, we have $F_{el} = F/\sin \alpha = F \cdot 2l/r = kq^2 \cdot 2l/r^3 \simeq 3.5 \times 10^{-3}$ N.

The tension in the thread is approximately equal to 3.5×10^{-3} N.

3. Three small charged bodies are on the same straight line. The body with a charge $q_2 = 1.0 \times 10^{-8}$ C is between the body with a charge $q_1 = -3.0 \times 10^{-8}$ C (at
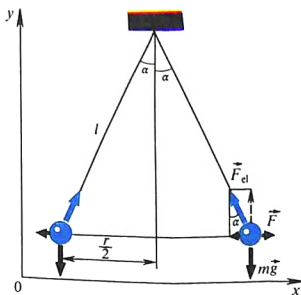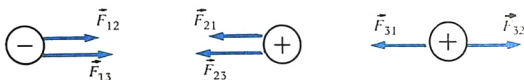

Fig. 96

Fig. 97

a distance $r_{12} = 0.4$ m from it) and the body with a charge $q_3 = 5.0 \times 10^{-8}$ C (at a distance $r_{23} = 0.6$ m). Find the resulting Coulomb forces applied to each body.

Solution. According to Coulomb's law the first and second bodies attract each other with a force

$$F_{12} = k\frac{|q_1||q_2|}{r_{12}^2}.$$

The second and third bodies repel each other with a force whose magnitude is given by

$$F_{23} = k\frac{|q_2||q_3|}{r_{23}^2}.$$

The magnitude of the force of attraction between the first and third bodies is

$$F_{31} = k\frac{|q_1||q_3|}{(r_{12} + r_{23})^2}.$$

Let us consider Coulomb's forces acting on the first body (Fig. 97). They have the same direction. Consequently, the magnitude of the resulting force is equal to the sum of the magnitudes of these forces:

$$F_1 = F_{12} + F_{13} = k|q_1|\left(\frac{|q_2|}{r_{12}^2} + \frac{|q_3|}{r_{13}^2}\right) \simeq 30 \times 10^{-6} \text{ N}.$$

Coulomb's forces applied to the second body also have the same direction (Fig. 96), and hence the magnitude of the resultant force is

$$F_2 = F_{12} + F_{23} = k|q_2|\left(\frac{|q_1|}{r_{12}^2} + \frac{|q_3|}{r_{23}^2}\right) \simeq 2.9 \times 10^{-6} \text{ N}.$$

The forces applied to the third body have opposite directions (Fig. 97). The force of attraction F_{31} exerted by the first body is stronger than the repulsive force F_{32} exerted by the second body. The magnitude of the resultant force is equal to the difference of the forces. The resultant force is directed along the larger force:

$$F_3 = F_{13} - F_{23} = k|q_3|\left[\frac{|q_1|}{(r_{12} + r_{23})^2} - \frac{|q_2|}{r_{23}^2}\right] \simeq 1.1 \times 10^{-6} \text{ N}.$$

Exercise 8

1. A glass rod rubbed against silk is known to acquire a positive charge. Determine experimentally the sign of the charge of a plastic pen rubbed against wool.

124

2. Bodies in close contact without friction are electrostatically charged, although insignificantly. Does this mean that two bodies can be charged without doing work?

3. Determine the force of interaction between the electron and the nucleus in a hydrogen atom if the distance between them is 0.5×10^{-8} cm.

4. What is the ratio of the Coulomb force of interaction between the electron and the nucleus in a hydrogen atom to the force of their gravitational interaction? The masses of the electron and the proton are $m_e = 9.11 \times 10^{-31}$ kg and $m_p = 1.67 \times 10^{-27}$ kg, and the gravitational constant $G = 6.67 \times 10^{-11}$ N·m²/kg².

5*. What would be the force of interaction between two water drops at a distance of 1 km, if $1^o/_o$ of the electrons contained in one of the drops whose mass is 0.03 g could be transferred to the other drop?

6. Two identical balls are at a distance of 40 cm from each other. The charges of the balls are 9×10^{-9} C and -2×10^{-9} C. The balls are brought in contact and then moved apart to the same distance. Find the forces of their interaction before and after the contact.

7. Two point charges of 1.0×10^{-8} C and 2.0×10^{-8} C are fixed at a distance of 1 m from each other in vacuum. A small body bearing a charge of -3×10^{-9} C is placed at the midpoint of the straight line connecting the charges. Find the magnitude and direction of the force acting on the body.

8.6. Short-range and Long-range Action

SHORT-RANGE ACTION. The law of interaction of fixed electric charges was established experimentally. However, the way of realization of this interaction remained unclear.

If we observe the action of one body on another located at a certain distance from it, before admitting that this action is direct, we are inclined to find out whether there exists a material link (threads, rods, etc.) between the bodies. If such links do exist, we explain the action of one body on the other with the help of these intermediate links. For example, when the driver of an old-fashioned bus turns the handle to open the door, successive segments of the connecting rod are compressed and start to move until the door is opened. In modern buses, the driver opens the door by sending compressed air through the tubes to the cylinder which controls the mechanism of the door. This can also be done with the help of an electromagnet by passing an electric current through its wires.

There is something in common in the three ways of opening the door: there exists a continuous connecting line between the driver and the door, a certain physical process occurring at each point of this line. The action is transmitted with the help of this process propagating from point to point at a certain rate and not instantaneously.

Thus, the action between the bodies at a certain distance can be explained in many cases by the presence of intermediate links transmitting the action. It would be reasonable to assume the presence of certain intermediate links if there is no medium or mediator between interacting bodies. Otherwise, we

shall have to assume that a body acts at the sites where it is absent.

Those who know nothing about the properties of air may think that the mouth or vocal chords of a speaking person produce a direct action on the ears and consider the transfer of sound by an invisible medium to be completely incomprehensible. However, the entire process of propagation of sound waves can be traced and their velocity can be calculated.

The assumption about intermediate links (or media) through which the interaction between bodies at a distance from each other is always transmitted from point to point constitutes the essence of the short-range action theory.

Many scientists supporting the theory of short-range action tried to explain the origin of gravitational or electromagnetic forces with the help of invisible flows surrounding planets and magnets or by unseen atmospheres around electrically charged bodies. These arguments had to be rejected for one reason or another.

LONG-RANGE ACTION. The short-range action theory was widely accepted until Newton established the law of universal gravitation, albeit without providing any explanation for this law. Subsequent advances in the study of the solar system had such a strong effect on the scientists that most of them became aware that there is no point in seeking mediators for transmitting interaction from one body to another.

Thus the theory of direct action at a distance through a free space appeared. According to this theory, an action is transferred instantaneously over as large distances as desired. Bodies are capable of "perceiving" the presence of other bodies without any medium connecting them. Those favouring this theory were not confused at the thought that a body produces an action where it is not present. "Don't we see a magnet or a charged rod attracting bodies through the empty space?" they said. The force of attraction does not change noticeably if we wrap, for example, a magnet in paper or put it into a wooden box. Moreover, even when it seems to us that the interaction is due to a direct contact between bodies, actually it is not so. When bodies or parts of the same body are in contact, there are always small gaps between them. For example, a load suspended by a thread does not tear it although there is nothing between atoms constituting it (Fig. 98). According to this theory, the long-range action is the only mode of action encountered everywhere.
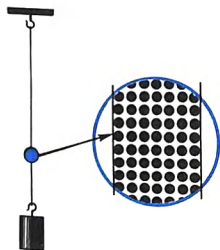

Fig. 98

126

The arguments against the short-range theory were quite strong. Besides, these arguments were confirmed by the achievements of Coulomb and Ampère who discovered the laws of interaction of charges and currents and were firm supporters of the long-range action theory.

If the evolution of science followed a straight-line path, it would seem that the victory of the long-range theory was ensured. As a matter of fact, however, the evolution of science rather takes place along a helical line. Completing a turn, science returns almost to the same concepts, although at a higher level. This was the case with the molecular kinetic theory. The Democritus atomistic hypothesis was rejected at one time by most of scientists. Then it was revived in a strict mathematical form and was confirmed in experiments. The same fate was met by the short-range theory.

Advances in the investigation of the laws of interaction of electric currents and charges were not inseparably linked with the idea of long-range action. In fact, the experimental investigation of forces themselves does not provide any idea about the way in which these forces are transferred. The primary goal was to obtain the mathematical expression for the forces, while their nature could be elucidated later.

8.7. Electric Field

FARADAY'S IDEAS. The final return to the short-range concepts was made by the English scientist M. Faraday and was completed by J. C. Maxwell.

According to the long-range theory, a charge directly perceives the presence of another charge. When charge A is displaced (Fig. 99), the force acting on charge B instantaneously changes its magnitude. In this case, neither charge B itself, nor the space surrounding it experience any change.

According to Faraday's idea, *electric charges do not act on each other directly. Each charge creates an electric field in the surrounding space. The field of one charge acts on another charge, and vice versa.* The field decreases as we move away from the charge creating it.

Initially, this idea just expressed Faraday's conviction that the action of one body on another through free space is impossible.

There was no proof of the existence of the field. Such a proof could not be obtained by studying the interaction of fixed charges. The short-range theory came into prominence after the investigation of electromagnetic interaction of moving charged particles. First the existence of time-dependent fields was proved, and only after this the conclusion about the presence of the electric field of fixed charges was drawn.

THE VELOCITY OF PROPAGATION OF ELECTROMAGNETIC INTERACTIONS. On the basis of Faraday's ideas, Maxwell proved theoretically that electromagnetic interactions should propagate in space at a finite velocity.

Fig. 99

This means that if we slightly displace charge A (Fig. 99), the force acting on charge B will change not instantaneously but in a time

$$t = |AB|/c, \tag{8.8}$$

where $|AB|$ is the distance between the charges and c is the velocity of propagation of electromagnetic interactions. Maxwell found that this velocity is equal to the velocity of light in vacuum, i.e. 300 000 km/s. Upon a displacement of charge A, the electric field around charge B changes after time t.

Thus, a certain process occurs between charges in vacuum as a result of which their interaction propagates at a finite velocity.

THE EXISTENCE OF A CERTAIN PROCESS IN THE SPACE BETWEEN INTERACTING BODIES LASTING FOR A FINITE TIME IS THE PRINCIPAL DIFFERENCE BETWEEN THE SHORT-RANGE AND LONG-RANGE THEORIES. All other arguments in favour of any of these theories cannot be considered decisive. True, it is difficult to carry out an experiment to verify equation (8.8) for the displacement of charges because of a very large value of the velocity c. However, it was no longer necessary after the invention of radio.

RADIO WAVES. The transfer of information with the help of electromagnetic waves is known as radio communication. Now we can simply read in newspapers that radio waves from the space station approaching Venus reach the Earth in more than four minutes. The station may be burnt in the atmosphere of the planet, but radio waves sent from it will travel in space for a long time. Thus, the electromagnetic field can be detected as something existing in reality.

WHAT IS ELECTRIC FIELD? We know that the electric field does exist in reality. Its properties can be studied experimentally. But we cannot determine its constituents. Here we reach the limit of the present-day knowledge.

A building consists of bricks, slabs, etc. The latter consist of molecules which in turn consist of atoms. Atoms are composed of elementary particles.

128

No simpler objects than elementary particles are known (at least for the time being). The same is true for the electric field. We do not know anything more elementary than the field. Therefore, we can say only the following about the origin of the electric field:

first, the field is material: it exists independently of us and of our knowledge about it;

second, it possesses certain properties which distinguish it from all other things in the world surrounding us. The establishment of these properties just forms our concepts about the electric field.

While studying the electric field, we come across a special kind of matter whose motion does not obey Newtonian mechanics. The discovery of the electric field led for the first time in the history of science to the profound idea about *the existence of different forms of matter, having their own laws.*

BASIC PROPERTIES OF ELECTRIC FIELD. The main property of the electric field is its action on electric charges with a certain force. From the action on the charge, the existence of the field and its distribution in space are established and all its characteristics are studied.

The electric field of fixed charges is called the ELECTROSTATIC FIELD. It does not change with time. *An electrostatic field is produced only by electric charges.* It exists in the space surrounding the charges and is inseparably connected with them.

In the course of electrodynamics, we shall study new properties of the field. We shall get acquainted with the time-dependent electric field which is no longer associated inseparably with charges. Static and varying fields have many properties in common. However, there are considerable differences between them. While studying the properties of the field, we shall refer to it as the electric field if a given property is equally inherent in static as well as varying fields.

8.8. Electric Field Strength. Principle of Superposition of Fields

ELECTRIC FIELD STRENGTH. An electric field is detected from the forces acting on a charge. We can state that we know everything what we need to know about a field if we know the force acting on any charge at any point of the field.

Therefore, we must introduce a characteristic of the field which will allow us to determine this force.

If we successively place at the same point in a field small charged bodies and measure the forces acting on them, it will turn out that the force exerted by the field on a charge is proportional to this charge. Indeed, let a field be created by a point charge q_1. According to Coulomb's law (8.2), the force acting on a charge q_2 is proportional to this charge. Therefore, *the ratio of the force acting on the charge placed at a given point of the field to the magnitude of this charge does not depend on the charge and can be regarded as a characteristic of the field at each point.* This characteristic is known as the ELECTRIC FIELD STRENGTH. Like force, electric field strength is a vector quantity. It is denoted by $\vec{E}$. If we denote the charge placed in the field by
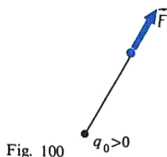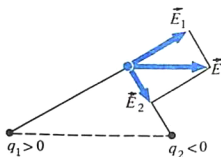
Fig. 100 $q_0 > 0$ $q_0 < 0$ Fig. 101 $q_1 > 0$ $q_2 < 0$

q instead of q_2, then, by definition, the electric field strength is

$$\vec{E} = \vec{F}/q. \tag{8.9}$$

The electric field strength is equal to the ratio of the force exerted by the field on a point charge to the magnitude of this charge.

Hence the force exerted by the field on a charge q is given by

$$\vec{F} = q\vec{E}. \tag{8.10}$$

The direction of vector $\vec{E}$ coincides with the direction of the force acting on a positive charge and is opposite to the direction of the force acting on a negative charge.

According to formula (8.9), the field strength can be expressed in SI units in newtons per coulomb (N/C).

FIELD STRENGTH OF A POINT CHARGE. Let us find the strength of the electric field created by a point charge q_0. According to Coulomb's law, this charge will act on another charge q with a force

$$F = k\frac{|q||q_0|}{r^2}.$$

The magnitude of the field strength of the point cha   .t a distance r from it is given by

$$E = \frac{F}{|q|} = k|q_0|/r^2. \tag{8.11}$$

At any point of the electric field, the vector of the field strength is directed along the straight line connecting this point and the charge, away from the charge if $q_0 > 0$ and towards the charge if $q_0 < 0$ (Fig. 100).

THE PRINCIPLE OF SUPERPOSITION OF FIELDS. If several forces act on a body, then in accordance with the laws of mechanics, the resultant force is equal to the geometrical sum of the forces: $\vec{F} = \vec{F}_1 + \vec{F}_2 + \dots$.

Electric charges experience the action of forces exerted by an electric field. If upon a superposition of the fields of several charges these fields do not affect one another, the resultant force exerted by all the fields must be equal to the geometrical sum of the forces exerted by each field. This is what happens in actual practice. Thus, the field strengths are added geometrically.

This is the essence of the PRINCIPLE OF SUPERPOSITION OF FIELDS[1]. This

---

[1] Superposition means overlapping.

principle is formulated as follows: if various charged particles produce electric fields of strengths $\vec{E}_1$, $\vec{E}_2$, $\vec{E}_3$, ... at a given point, the resultant field strength at this point is given by

$$\vec{E} = \vec{E}_1 + \vec{E}_2 + \vec{E}_3 + \dots \tag{8.12}$$

In view of the superposition principle, to determine the field strength of a system of charged particles of any point, it is sufficient to know expression (8.11) for the field strength of a point charge. Figure 101 illustrates how the strength $\vec{E}$ of the field produced by two point charges q_1 and q_2 is determined.

The introduction of the idea of the electric field allows us to break the problem of calculating the interaction forces between charged particles into two parts. First the field strength due to the charges is calculated, and then forces are determined from the known field strength. Such a splitting of the problem into parts usually simplifies the calculation of forces.

8.9. Lines of Force of Electric Field

We can get an idea about the field distribution by drawing field strength vectors at several points in space (Fig. 102). The pattern will be more visual if we draw continuous lines such that their tangents at each point through which they pass coincide with the electric field strength vector. These lines are known as LINES OF FORCE of the electric field, or electric field lines (Fig. 103).

It would be wrong to think (like Faraday) that electric field lines are the real formations similar to stretched elastic strings or cords. They just help to visually represent the field distribution in space and are not more real than meridians and parallels on the globe.

Lines of force, however, can be made "visible". If we thoroughly mix oblong crystals of an insulator (say, quinine), with a viscous liquid (like castor oil) and place charged bodies in this medium, the crystals "align" themselves in chains along the field lines.

Figures 104-107 represent field lines for different fields: of a positively charged ball (Fig. 104); of two oppositely charged balls (Fig. 105); of two like-charged balls (Fig. 106), and of two plates bearing equal and opposite charges (Fig. 107). The latter example is of special importance. Figure 107 shows that in the space between the plates, away from their edges, the field lines are parallel: the electric field is the same at all points in this region.

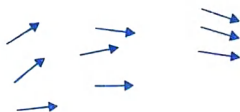*An electric field whose strength is the same at all points in space is called*
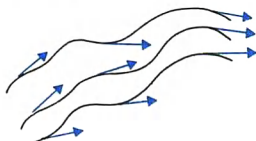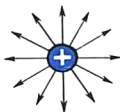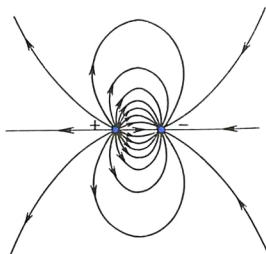

Fig. 102

Fig. 103

Fig. 104

Fig. 105

uniform. An electric field can be assumed nearly uniform in a bounded region of space if the field strength in this region varies insignificantly.

The electric field lines are not closed. They originate at positive charges and terminate at negative charges[1]. The lines are continuous and do not intersect, since their intersection would mean the absence of a certain direction of the electric field strength at a given point. The lines of force originate or terminate at charged bodies and then diverge in different directions (Fig. 104). Hence the *density of the field lines is higher near charged bodies where the field strength is also higher.*

?   1. What is the difference between the short-range and the long-range theories?
2. List the main properties of the electrostatic field.
3. Give the definition of electric field strength.
4. What is the field strength of a point charge equal to?
5. Formulate the principle of superposition.
6. What are the lines of force of electric field?
7. Draw the lines of force for a uniform electric field.

---

[1] The field lines shown in Fig. 107 also terminate at negative charges situated far away.
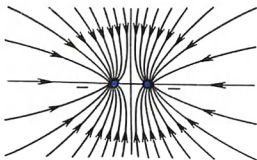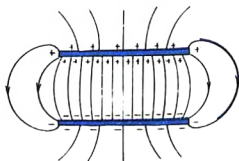

Fig. 106

Fig. 107

8.10. Conductors in an Electrostatic Field

FREE CHARGES. Conductors (which first of all include metals) contain charged particles capable of moving freely in a conductor under the action of an electric field. For this reason, the charges of these particles are called FREE CHARGES.

The carriers of free charges in metals are electrons. When a metal consists of neutral atoms, the latter start interacting. Due to this interaction, the electrons of the outer shells of the atoms completely lose the bonds with "their own" atoms and become the "property" of the entire conductor as a whole. As a result, the positively charged ions turn out to be surrounded by a negatively charged "gas", formed by collectivized electrons (Fig. 108). Free electrons take part in thermal motion and can move over a piece of metal in any direction.

ELECTROSTATIC FIELD INSIDE A CONDUCTOR. Due to the presence of free charges in a conductor, there is no electrostatic field inside it. If the electric field strength were different from zero, the field would cause the ordered motion of free charges, i.e. an electric current would flow in the conductor. The statement about the absence of electric field in a conductor is equally valid both for a charged conductor and for a neutral conductor placed in an external electrostatic field.[1]

Let us consider by way of an example a neutral plate introduced into a uniform field (Fig. 109) and analyze the process as a result of which the field strength inside the conductor turns out to be zero. Under the action of the electric field, the electrons of the plate start to move from right to left. At the initial moment (when the conductor is just introduced into the field) an electric current appears. The left part of the plate is charged negatively, and the right part acquires a positive charge. This is the manifestation of ELECTROSTATIC INDUCTION. (If we divide the plate into two parts along the line *MN*, the two parts turn out to be charged.) The resulting charges produce their own field (the lines of force of this field are shown in Fig. 110 by the dashed lines), which is superimposed on the external field and balances it. The charges are redistributed in an extremely short time so that the strength of the resultant field inside the plate becomes zero, and the motion of charges ceases. Otherwise, an electric current would continuously flow in the conductor and heat would be liberated all the time. However, this is impossible according to the energy conservation law.

---

[1] Naturally, individual charged particles (electrons and ions) produce microscopic fields. These fields, however, are mutually compensated, and the average value of the field strength turns out to be zero.
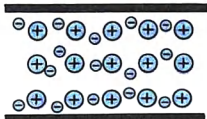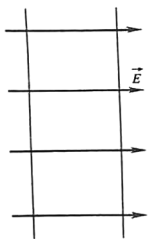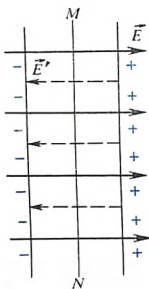

Fig. 108
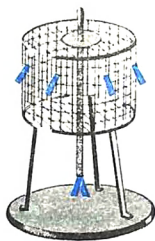
Fig. 109   Fig. 110   Fig. 111

Thus, there is no electrostatic field inside a conductor. This fact is used in the so-called ELECTROSTATIC SHIELDING. In order to protect the instruments sensitive to the electric field, they are enclosed in metal boxes.

Outside a conductor, the lines of force of an electrostatic field are normal to its surface. Otherwise, there would be a component of the field strength along the surface of the conductor, and an electric current would flow over its surface.

ELECTRIC CHARGE OF CONDUCTORS. When the charges are in equilibrium, not only the field strength inside a conductor is zero, but its charge is equal to zero as well. *The entire electrostatic charge of the conductor is concentrated on its surface.* Indeed, if there were a charge inside the conductor, it would produce a field in the surrounding space. But there is no field in the conductor. Consequently, the charges can be arranged only on the surface of the conductor.

This conclusion is valid both for neutral conductors in an electric field and for charged conductors.

The absence of a charge inside a conductor can be confirmed with the help of simple experiments, for example, by an experiment with a cylinder made of a wire netting (Fig. 111). Light strips of tin foil are attached to the surface of the cylinder. Two more strips are fixed to a conducting movable rod inside the cylinder. If we supply to the cylinder a charge from say, a frictional machine, the strips diverge through a certain angle since the charge supplied to them is repelled by the like charge of the cylinder or of a neighbouring strip. If, however, we introduce the strips on the rod inside the cylinder, they do not diverge since their charge is zero.

8.11. Electric Field Strength of a Uniformly Charged Conducting Sphere and an Infinite Plane

Let us determine the strengths of electric fields produced by charged bodies of a simple shape, viz. a sphere and a plane. Many bodies in nature and in engineering have a nearly spherical shape:

atomic nuclei, rain drops, planets, etc. Plane surfaces are also encountered frequently. Besides, a small region of any surface can be regarded as nearly plane.

FIELD OF A SPHERE. Let us consider a charged conducting sphere of radius R. A charge q is uniformly distributed over the surface of the sphere. According to symmetry considerations, the lines of force of the electric field are directed along the continuations of the radii of the sphere (Fig. 112).

Pay attention to the fact that the lines of force outside the sphere are distributed in space in the same way as the field lines of a point charge (Fig. 113). If the patterns of the fields coincide, it can be expected that the field strengths are identical. Therefore, the field strength at a distance $r \geqslant R$ from the centre of the sphere is determined by the same formula (8.11) as the field strength of a point charge placed at the centre of the sphere:

$$E = \frac{k|q|}{r^2}. \tag{8.13}$$

Strict calculation leads to the same result.

Inside a conducting sphere ($r < R$), the electric field strength is zero.

FIELD OF A PLANE. The distribution of electric charge over the surface of a charged body is characterized by a special quantity, viz. the surface density σ of the charge. The surface charge density is the ratio of the charge to the area of the surface over which it is distributed. If a charge q is distributed uniformly over a surface whose area is S, we can write

$$\sigma = q/S.$$

The SI unit of the surface charge density is coulomb per squared metre (C/m^2).

It follows from symmetry considerations that the electric filed lines cannot be inclined to the plane in any direction. All directions are equivalent, and there are no grounds for inclination. Therefore, the lines of force of the electric field produced by an infinite uniformly charged plane are straight lines normal to the plane (Fig. 114). The field of the infinite plane is uniform, i. e. the field strength is the same at all points of space irrespective of the distance from the plane. It is determined by the surface density σ of the charge.

In order to determine the relation connecting the surface charge density σ and the field strength, we can employ a method widely used in physics and

based on the knowledge of dimensions of physical quantities. The dimensions of the electric field strength are N/C, while the surface charge density is measured in C/m².

To obtain the correct dimensions of the field strength in this case, we must assume that the field strength is proportional to the product of the coefficient k (see (8.11)) and the surface charge density σ, i.e. $E \propto k\sigma$. The coefficient k is expressed in $(N \cdot m^2)/C^2$. Then the dimensions of the field strength are $(N \cdot m^2/C^2) \cdot (C/m^2) = N/C$.

The proportionality factor can be found only with the help of calculations, which lead to the following result:

$$E = k\,2\pi\,|\,\sigma\,|. \tag{8.14}$$

Infinite planes do not exist. If, however, the size of a plane is large in comparison with the distance from the plane to a given point, the field strength component normal to the plane at this point is practically the same as the field strength of an infinite plane.

8.12. Dielectrics in Electrostatic Field. Two Types of Dielectrics

An insulator, or dielectric[1], consists of atoms or molecules which are neutral in the whole. In a neutral atom, electric charges (to be more precise, electrically charged particles, viz. electrons and nuclei) are BOUND to one another and unlike free electrons of a conductor cannot move freely over the entire volume of the substance under the action of the field.

The difference in the structure of conductors and dielectrics determines their different behaviour in an electrostatic field. The electrostatic field can exist within a dielectric, the latter producing a certain effect on the field.

With the help of a simple experiment, we can show that a neutral dielectric can produce an electric field. Figure 115 shows a charged electrometer with a metallic disc at the end of the rod. If a neutral dielectric is brought to the disc (say, a thick piece of glass), the pointer of the electrometer comes closer to the rod (Fig. 116). This can occur only if the dielectric brought in the field of the charged disc produces an electric field itself. This field affects the charge distribution in the electrometer rod, reducing the charge of the pointer and the rod and increasing accordingly the charge of the disc.

ELECTRIC PROPERTIES OF NEUTRAL ATOMS AND MOLECULES. In order to explain how a neutral dielectric produces an electric field, we must first study the electric properties of neutral atoms and molecules.

Atoms and molecules consist of positively charged particles, viz. nuclei, and negatively charged particles, viz. electrons. Figure 117 represents

[1] Insulators are called in physics dielectrics (from the Greek *dia* meaning "through"). The term dielectric is applied to substances through which electromagnetic interactions are transferred.
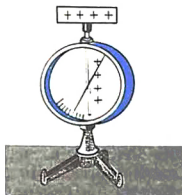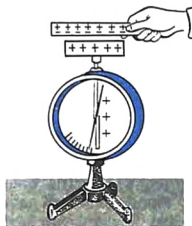
Fig. 115   Fig. 116

a schematic diagram of the simplest atom – the hydrogen atom. The positive charge of the atom, i.e. the charge of its nucleus, is concentrated at its centre. The electron moves in the atom at a high velocity. It completes a turn around the nucleus in a very short time of the order of 10^{-15} s. Therefore, it makes a million turns during 10^{-9} s, and hence it manages to pass through any points *1* and *2* arranged symmetrically about the nucleus a million times during this time interval. This circumstance leads to the assumption that the centre of distribution of the negative charge, averaged over time, is at the centre of the atom, i.e. it coincides with the positively charged nucleus.

This, however, is not always the case. Let us consider a molecule of common salt NaCl. The outer shell of a sodium atom has a single valence electron which is weakly bound to the atom. Chlorine has seven valence electrons. When the NaCl molecule is formed, the single valence electron of sodium is captured by the chlorine atom. The two neutral atoms form a system of two ions having unlike charges (Fig. 118). Now the positive and negative charges are not distributed symmetrically over the volume of the molecule: the centre of distribution of the positive charge lies in the sodium ion and of the negative charge in the chlorine ion.

ELECTRIC DIPOLE. At a large distance from the molecule, it can be considered approximately as a system of two point charges which are equal in magnitude and opposite in sign, and are separated by a certain distance *l* (Fig. 119). Such a system is neutral as a whole and is called an ELECTRIC DIPOLE.

TWO TYPES OF DIELECTRICS. Dielectrics can be divided into two categories:

POLAR dielectrics consisting of molecules in which the centres of distribution of positive and negative charges do not coincide, and
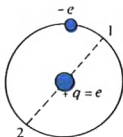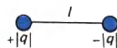

Fig. 117

Fig. 118

Fig. 119

137

NONPOLAR dielectrics consisting of atoms or molecules in which the centres of distribution of positive and negative charges coincide. Polar dielectrics include alcohols, water, etc. Nonpolar dielectrics are inert gases, oxygen, hydrogen, benzene, polyethylene, etc.

8.13. Polarization of Dielectrics. Permittivity

POLARIZATION OF POLAR DIELECTRICS. A polar dielectric consists of molecules that can be regarded as electric dipoles. Thermal motion leads to a random orientation of dipoles (Fig. 120a). For this reason, the electric charge on the surface of a dielectric, as well as the charge of any part of its volume containing a large number of molecules (hatched rectangle in Fig. 120a), is equal to zero on the average. The electric field strength in a dielectric on the average is also equal to zero.

Let us place a dielectric between two parallel metallic plates bearing unlike charges. If the size of the plates is much larger than their separation, the field between the plates is uniform (see Sec. 8.9). This field will exert on the dipole two equal and opposite forces (Fig. 121). These forces create a torque which tends to rotate the dipole so that its axis becomes parallel to the field lines (Fig. 122). The positive charges are displaced in the direction of the field and the negative charges in the opposite direction.

The displacement of positive and negative bound charges of a dielectric in the opposite directions is called polarization.

Random thermal motion hinders the ordered orientation of all the dipoles in a dielectric. Only at absolute zero would all the dipoles be aligned along the lines of force. Thus, under the action of the field only a partial orientation of electric dipoles takes place. This means that on the average, the number of dipoles oriented along the field is greater than the number of dipoles oriented against the field. Figure 120b shows that the negative charges of dipoles appear predominantly on the surface of the dielectric at the positive plate, and positive charges at the negative plate. As a result, a bound charge appears on the surface of the dielectric. Inside the dielectric the positive and negative charges of dipoles compensate each other and the mean bound electric charge is, as before, equal to zero.

POLARIZATION OF NONPOLAR DIELECTRICS. A nonpolar dielectric is also polarized in an electric field. Under the action of the field,
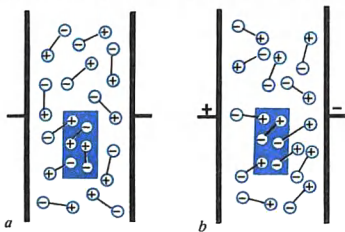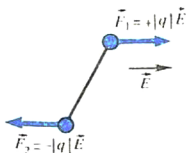

Fig. 120   a   b

$$\vec{F}_1 = +|q|\vec{E}$$

$$\vec{E}$$

$$\vec{F}_2 = -|q|\vec{E}$$

Fig. 121

$$-|q| \qquad +|q|$$

$$\vec{E}$$
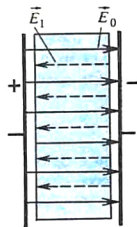
Fig. 122

$$\vec{E}_1 \qquad \vec{E}_0$$

Fig. 123

the positive and negative charges of molecules are displaced in opposite directions, and the centres of distribution of positive and negative charges do not coincide any longer, as in the case of a polar molecule. Such deformed molecules can be treated as electric dipoles whose axes are oriented along the field. Bound charges appear on the surfaces of the dielectric facing the charged plates as in the case of polarization of a polar dielectric.

The forces of attraction between neighbouring molecules of a dielectric are much weaker than the forces binding unlike charges in a molecule. For this reason, the molecules of a dielectric do not break when the sample is cut or torn but remain intact on the surface of the cut. As a result, the parts of a polarized dielectric are electrically neutral.

PERMITTIVITY. A bound charge produces an electric field of strength $\vec{E}_1$ directed in the dielectric against the external field of the charges on the plates (Fig. 123). Thus, the field in the dielectric is weakened. The degree of weakening of the field depends on the properties of the dielectric.

The electric properties of dielectrics are characterized by a special quantity known as DIELECTRIC PERMITTIVITY.

The permittivity of a medium is a physical quantity equal to the ratio of the magnitude of the field strength E_0 in vacuum to the magnitude of the electric field strength E in a uniform dielectric.

Denoting the permittivity by ε, we can write $\varepsilon = E_0/E$[1].

Accordingly, the electric field of a point charge, a charged sphere and a plane placed in a uniform dielectric, is weaker than in vacuum by a factor of ε. Instead of formula (8.13), we must use the following formula for a point charge and a sphere:

$$E = k\frac{|q|}{\varepsilon r^2}. \tag{8.15}$$

For a charged plane in a uniform dielectric, we must take instead of (8.14)

$$E = k\frac{2\pi|\sigma|}{\varepsilon}. \tag{8.16}$$

---

[1] This formula is valid only for a homogeneous medium filling the entire space, or in case of a body having a special symmetry, like a uniform plate in a capacitor. For a body of an arbitrary shape, the dependence between E and E_0 is much more complicated and is determined by the shape of the body and its orientation relative to E_0.

The force of interaction of point charges in a uniform dielectric is also smaller by a factor of ε due to a decrease in the field strength. Coulomb's law for charges in a uniform dielectric has the form

$$F = k \frac{|q_1||q_2|}{\varepsilon r^2}.$$ (8.17)

Thus, unlike the forces of universal gravitation, the forces of interaction between charged bodies depend on the properties of the medium in which they are placed.

?

1. What is the value of the field strength inside a conductor if its charges are in equilibrium?
2. In which way is the charge supplied to a conductor distributed over it?
3. What is the strength of the field of a charged conducting sphere equal to?
4. Write the expression for the field strength of a uniformly charged plane.
5. Which dielectrics are known as polar and nonpolar?
6. Which phenomenon is called the polarization of a dielectric?
7. Give the definition of dielectric permittivity.

8.14. Potential Energy of a Charged Body in a Uniform Electrostatic Field

ELECTRIC ENERGY. Charged bodies attract or repel one another. When charged bodies (for example, the strips of an electroscope) are displaced, the forces acting among them do work. It is known from mechanics that a system capable of doing work due to the interaction among bodies possesses a potential energy. Hence, a SYSTEM OF CHARGED BODIES HAS A POTENTIAL ENERGY called ELECTROSTATIC or ELECTRIC ENERGY.

The energy of interaction of electrons with the nucleus in an atom and the energy of interatomic interaction in a molecule (chemical energy) is mainly the electric energy. A huge amount of electric energy is stored in an atomic nucleus. During the operation of the nuclear reactor of a power plant, heat is liberated just at the expense of this energy.

From the point of view of the short-range theory, the electric field produced by other charges directly acts on a charge. When the charge is displaced, the forces exerted on it by the field do work. (Henceforth, we shall speak simply of the work done by the field for the sake of brevity). Thus, we can state that a charged body in an electric field possesses energy. Let us find the potential energy of a charge in a uniform electric field.

WORK DONE UPON A DISPLACEMENT OF A CHARGE IN A UNIFORM ELECTROSTATIC FIELD. A uniform field is produced, for example, by large metallic plates bearing unlike charges. Such a field exerts on a charge a constant force $\vec{F} = q\vec{E}$ in the same way as the Earth acts on a stone raised above its surface with a constant force $\vec{F} = m\vec{g}$. Suppose that
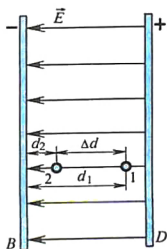
Fig. 124

the plates are in the vertical position (Fig. 124): the left plate B has a negative charge and the right plate D has a positive charge. Let us calculate the work done by the field upon the displacement of a positive charge q from point 1 at a distance d_1 from plate B to point 2 at a distance $d_2 < d_1$ from the same plate. Points 1 and 2 are on the same line of force.

The electric field does a positive work over the path $\Delta d = d_1 - d_2$:

$$A = qE(d_1 - d_2) = -(qEd_2 - qEd_1).\qquad(8.18)$$

This work does not depend on the shape of the path just as the work done by the force of gravity does not depend on the trajectory.

The proof of this law was given in "Senior physics 1", and we shall not repeat it for the constant force $q\vec{E}$.

POTENTIAL ENERGY. *If the work does not depend on the shape of the path, it is equal to the change in potential energy (taken with the minus sign):*

$$A = -(W_{p2} - W_{p1}) = -\Delta W_p.\qquad(8.19)$$

This question was discussed in detail in "Senior Physics 1".

Comparing the expression (8.18) with the general definition (8.19) of the potential energy, we see that the potential energy of a charge in a uniform electrostatic field is

$$\underline{W_p = qEd.}\qquad(8.20)$$

This formula is similar to the formula $W_p = mgh$ for the potential energy of a body above the surface of the Earth. But unlike the mass, the charge q can be either positive or negative. If $q < 0$, the potential energy (8.20) is negative[1].

If a field does positive work, the potential energy of the charged body decreases: $\Delta W_p < 0$. In accordance with the law of energy conservation, its

---

[1] Besides, while the direction of the force of gravity cannot be changed by us, the direction of the electric field can be easily changed by reversing the signs of the charges on plates B and D. In this case, the minus sign will appear in the expression (8.20) for the potential energy.

kinetic energy simultaneously increases. This principle is used for accelerating electrons by an electric field in electronic and cathode-ray tubes. If, on the contrary, the work is negative (for example, when a positively charged particle moves in a direction opposite to the field strength $\vec{E}$; this motion is similar to the motion of a stone thrown upwards), then $\Delta W_p > 0$. The potential energy increases, while the kinetic energy decreases and the particle is decelerated.

When a charge returns to the initial point 1 along a closed path the work is equal to zero,

$$A = -\Delta W_p = -(W_{p2} - W_{p1}) = 0.$$

ZERO LEVEL OF POTENTIAL ENERGY. The potential energy (8.20) is equal to zero on the surface of plate B. This means that the zero level of potential energy coincides with plate B. However, as in the case of gravitational forces, the zero level of potential energy is chosen arbitrarily. We can assume that $W_p = 0$ at a distance d_2 from plate B. Then

$$W_p = qEd - qEd_2.$$

Physical meaning is attached not to the potential energy but to the difference in its values, which is determined by the work of the field upon the displacement of a charge from the initial to the final position.

8.15. Potential of Electrostatic Field and Potential Difference

POTENTIAL FIELDS. When a charged body is displaced from one point to another, the work of any electrostatic field does not depend on the shape of the path as is the case for a uniform field. The work of an electrostatic field on a closed path is always zero. The fields having this property are known as POTENTIAL FIELDS. In particular, the electrostatic field of a point charge is of potential nature.

The work of a potential field can be expressed in terms of the change in potential energy. The formula $A = -(W_{p2} - W_{p1})$ is valid for an arbitrary electrostatic field. However, the energy is expressed by formula (8.20) only in the case of a uniform field.

POTENTIAL. The potential energy of a charge in an electrostatic field is proportional to the charge. This is true for a uniform field (see formula (8.20)) as well as for any other field. Consequently, the *ratio of the potential energy to the charge does not depend on the charge placed in the field.*

This circumstance makes it possible to introduce a new quantitative characteristic of the field, viz. the POTENTIAL which does not depend on the charge placed in the field. The potential of an electrostatic field is the ratio of the potential energy of a charge in the field to this charge.

According to this definition, the potential is given by

$$\varphi = \frac{W_p}{q}. \tag{8.21}$$

The field strength $\vec{E}$ is a vector and the force characteristic of the field. It determines the force acting on charge q at a given point of the field. The potential φ is a scalar; it is the characteristic of the field from the energy viewpoint. The potential determines the potential energy of charge q at a given point of the field.

If we take the negative plate as the zero level of potential energy, and hence of the potential (Fig. 124), then according to formulas (8.20) and (8.21), the potential of a uniform field is

$$\varphi = \frac{W_p}{q} = Ed. \tag{8.22}$$

POTENTIAL DIFFERENCE. Like the potential energy the value of potential at a given point depends on the choice of the zero level of potential. Of practical importance is not the value of potential at a point but the change in potential, which does not depend on the choice of the zero level of potential.

Since the potential energy $W_p = q\,\varphi$, the work is given by

$$A = -(W_{p2} - W_{p1}) = -q(\varphi_2 - \varphi_1) = -q\,\Delta\varphi. \tag{8.23}$$

Henceforth, instead of the change in potential $\Delta\varphi = \varphi_2 - \varphi_1$ which is the difference in the values of potential at the final and initial point, we shall use another quantity, viz. the POTENTIAL DIFFERENCE. *The potential difference is the difference in the values of potential at the initial and final points of the path:*

$$U = \varphi_1 - \varphi_2 = -\Delta\varphi. \tag{8.24}$$

The potential difference is often called VOLTAGE. The potential variation differs from voltage only in sign.

It is more convenient to deal with the potential difference, or voltage U, than with the change in potential $\Delta\varphi$, especially when studying electric current.

According to formulas (8.23) and (8.24), the potential difference is given by

$$U = \varphi_1 - \varphi_2 = -\Delta\varphi = \frac{A}{q}. \tag{8.25}$$

The potential difference (voltage) between two points is equal to the ratio of the work done by the field in displacing a charge from the initial point to the final point to this charge.

Knowing the voltage of an electric lighting system, we can determine the work which can be done by the electric field to carry a unit charge from a socket terminal to any other point in the circuit. The concept of potential difference will be used throughout this course of physics.

UNIT OF POTENTIAL DIFFERENCE. The unit of potential difference is derived with the help of formula (8.25). In SI, the work is expressed in joules and charge in coulombs. Hence, the potential difference between two points is equal to unity if the electric field does a work of 1 J to carry a charge of 1 C from one point to another. This unit is known as the volt (V). $1\ V = 1\ J/1\ C.$

1. Which fields are called potential fields?
2. How is the change in potential energy related to work?
3. What is the expression for the potential energy of a particle in a uniform electric field?
4. Give the definition of potential.
5. What is the potential difference between two points in a field equal to?

8.16. Electrostatic Field Potential of a Point Charge

The derivation of the formula for electrostatic field of a point charge q as a function of distance r is quite complicated. The field strength of a point charge decreases with increasing distance. Consequently, to determine potential, we must calculate the work done by a varying Coulomb force.

The expression for the field potential of a point charge has the form

$$\varphi = k \frac{q}{\varepsilon r}. \tag{8.26}$$

The potentials of points in the field of a positive charge $(q > 0)$ are also positive $(\varphi > 0)$, and for a negative charge $(q < 0)$, negative $(\varphi < 0)$.

Formula (8.26) corresponds to a certain choice of the zero potential level. The potentials of points at infinite distances from a charge are assumed by convention to be zero: $\varphi \to 0$ as $r \to \infty$. Such a choice of zero potential level is convenient but not obligatory. We could add any constant quantity to potential (8.26). This would not change the potential difference between any two points, which is actually of practical importance.

If the potential of infinitely remote points is taken as zero, the potential of the field of a point charge acquires a simple physical meaning. Substituting into formula (8.25) the value $\varphi_2 = 0$, we obtain

$$\varphi_1 = \varphi(r) = \frac{A}{q}.$$

Consequently, the potential of the electrostatic field at a distance r from the point charge is numerically equal to the work done by the field in carrying a unit positive charge from a given point of space to an infinitely remote point.

Formula (8.26) is also valid for the potential of the field of a uniformly charged sphere at distances larger than or equal to its radius, since the field of a uniformly charged sphere coincides on its surface with the field of a point charge placed at the centre of the sphere.

We have considered the potential of the field of a point charge. The charge of any object can be mentally divided into small elements such that each of them can be treated as a point charge. Then the potential of the field at an arbitrary point will be determined as the algebraic sum of the potentials φ_1, φ_2, φ_3, etc. created by individual point charges:

$$\varphi = \varphi_1 + \varphi_2 + \varphi_3 + \ldots. \tag{8.27}$$

This relation follows from the principle of superposition of fields.

POTENTIAL ENERGY OF INTERACTION OF TWO POINT CHARGES. From the expression for the field potential of a point charge, we can calculate the potential energy of interaction of two point charges. In particular, this can be the energy of interaction of an electron with an atomic nucleus.

The potential energy of charge q_2 in the electric field of a point charge q_1 is equal to the product of charge q_2 and potential $\varphi(q_1)$ of the field produced by charge q_1: $W_p = q_2 \varphi(q_1)$.

Using formula (8.26) and replacing q by q_1, we obtain the required expression for the energy:

$$W_p = q_2 \varphi(q_1) = k \frac{q_1 q_2}{\varepsilon r}.$$

If charges q_1 and q_2 are of the same sign, the potential energy of their interaction is positive. The smaller the distance between the charges, the higher the potential energy of their interaction since the work that can be done by Coulomb forces upon repulsion of charges is larger. If the charges are unlike, the energy is negative, and its maximum value (equal to zero) is attained at $r \to \infty$. The larger the value of r, the larger the work done by attractive forces in bringing the charges closer.

8.17. Relation Between the Field Strength and Potential Difference of Electrostatic Field. Equipotential Surfaces

There exists a certain dependence between the electrostatic field strength and the potential difference. Let a charge q be displaced in the direction of a uniform electric field $\vec{E}$ from point 1 to point 2 at a distance Δd from point 1 (Fig. 125). The work done by the electric field is given by

$$A = qE \Delta d.$$

According to formula (8.25), this work can be expressed in terms of the potential difference between points 1 and 2:

$$A = q(\varphi_1 - \varphi_2) = qU. \tag{8.28}$$

Equating these expressions for work, we can find the magnitude of the electric field strength:
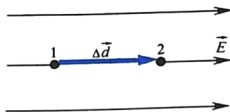
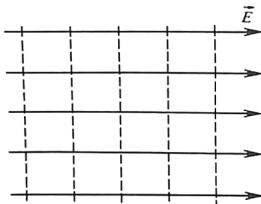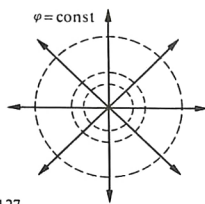$$E = \frac{U}{\Delta d}. \tag{8.29}$$

Fig. 125

Fig. 126

Fig. 127

In this formula, U is the potential difference between points 1 and 2 connected by the displacement vector $\Delta \vec{d}$, which coincides in direction with vector $\vec{E}$ (Fig. 125).

Formula (8.29) indicates that the smaller the change in the potential over the distance Δd, the smaller the strength of the electrostatic field. If the potential does not change at all, the field strength is zero.

Since the electrostatic field does positive work $A = q(\varphi_1 - \varphi_2)$ in displacing a positive charge in the direction of vector $\vec{E}$, the potential φ_1 is higher than the potential φ_2. Consequently, the *electric field strength is directed towards decreasing potential.*

In a small region of space, any electrostatic field can be considered to be uniform. Therefore, formula (8.29) is valid for an arbitrary electrostatic field if the distance Δd is so small that the change in the field strength over this distance can be neglected.

UNIT OF ELECTRIC FIELD STRENGTH. The SI unit of electric field strength is established with the help of formula (8.29). The electric field strength is equal to unity if the potential difference between two points separated by 1 m in a uniform field is 1 V. This unit has the dimensions of VOLT PER METRE (V/m).

It was mentioned above that the field strength can be expressed in newtons per coulomb. Indeed,

$$1 \text{ V/m} = 1 \text{ (J/C)} \cdot (1/\text{m}) = 1 \text{ (N} \cdot \text{m/C)} \cdot (1/\text{m}) = 1 \text{ N/C}.$$

EQUIPOTENTIAL SURFACES. If a charge is displaced at right angles to the lines of force, the electric field does no work since the force is normal to the displacement. Consequently, if we draw a surface perpendicular to the lines of force at each point, no work is done when a charge is moved over such a surface. This, in turn, means that *all points on the surface perpendicular to the lines of force have the same potential.* Surfaces of equal potential are known as equipotential surfaces.

The equipotential surfaces of a uniform field are planes (Fig. 126), while the equipotential surfaces of the field of a point charge are concentric spheres (Fig. 127). Equipotential surfaces of the field of a dipole are given on Fig. 128.

Like lines of force, equipotential surfaces qualitatively characterize the spatial distribution of a field. The electric field strength vector is normal to
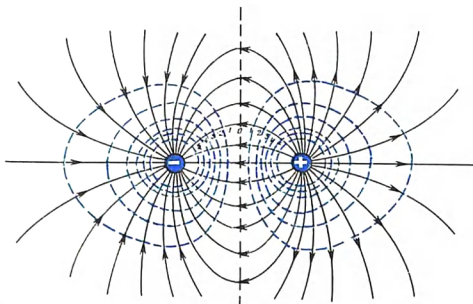
146

Fig. 128

equipotential surfaces and is directed towards decreasing potential. For example, the potential of the field of a positive point charge decreases as we move away from the charge, and the field strength is directed away from the charge along the radii of the concentric spheres (Fig. 127).

The surface of any conductor in an electrostatic field is an equipotential surface. Indeed, the lines of force are normal to the surface of the conductor. Moreover, not only the surface, but *all points inside the conductor are also at the same potential.* The field strength inside the conductor is zero, and hence the potential difference between any two points of the conductor is also zero.

8.18. Measurement of Potential Difference

The potential difference is measured with the help of an instrument called ELECTROMETER. Figure 129 represents an electrometer of a simple construction. Its main part is a light aluminium pointer fixed to a metallic rod with the help of a horizontal axle. The centre of gravity of the pointer is below the axle so that before measurements are taken, the pointer is in the vertical position. The rod and the pointer are enclosed into a metallic shell with an ebonite cork insulating it from the rod. The pointer is observed through a glass window. The electrometer resembles an electroscope, the only difference being that the former has a metallic shell.

In order to measure the potential difference between two conductors, one of them is connected to the rod of the electrometer and the other to its shell. (If the potential of a body relative to the Earth has to be measured, the body is connected to the rod with a wire, while the shell is earthed.) Then the potential difference between the rod and the shell is the same as the one to be measured. The electric field inside the electrometer depends only on this potential difference since no external electrostatic field of charged or polarized bodies penetrates the metallic shell of the instrument (electrostatic shielding). In turn, the forces acting on the pointer are uniquely determined by the field distribution inside the electrometer. To get an idea of the magni-

147

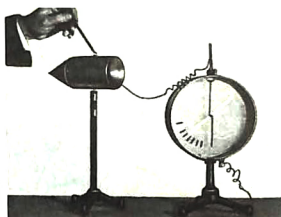Fig. 129   Fig. 130

tude of the potential difference from the position of the pointer, the instrument must be graduated. For this we must find out which angles of deflection of the pointer correspond to known values of the voltage between charged conductors.

Using an electrometer, we can easily prove by experiment that all points of a conductor have the same potential relative to the Earth. For this purpose, various parts of the conductor are connected consecutively by a wire to the rod of the electrometer whose shell is earthed (Fig. 130). The readings of the electrometer will not change.

?

1. What is the expression for the potential of the field produced by a point charge?
2. What is the relation between potential difference and electric field strength?
3. What is meant by a unit potential difference of 1 V?
4. What is the potential difference between two points of a charged conductor?

Problems with Solutions

While solving problems where the concept of electric field strength is used, you must be able to operate with formulas (8.10) and (8.15) which determine the force exerted on the charge by the electric field and the field strength of a point charge. The strength of the electric field produced by several point charges is equal to the geometrical sum of field strengths due to individual charges.

The work done by the forces exerted on the charge by the field is expressed in terms of the difference in potential energies of potential difference [formula (8.28)]. The potential of a uniform field is defined by formula (8.22), while the potential of the field produced by a point charge is given by formula (8.26).

While solving problems, you must bear in mind that all points of a conductor have the same potential and that the field strength inside it is zero.

148

1. Two identical positive point charges are at a distance r from each other in a homogeneous medium having a dielectric permittivity ε. Find the electric field strength and potential at a point lying at the same distance r from the two charges.

Solution. According to the principle of superposition, the required field strength $\vec{E}$ is equal to the geometrical sum of the field strengths due to each charge (Fig. 131). The magnitudes of the field strengths due to each charge are $E_1 = E_2 = k |q|/\varepsilon r^2$. The diagonal of the parallelogram plotted on vectors $\vec{E}_1$ and $\vec{E}_2$ is the strength of the resultant field, whose magnitude is given by

$$E = 2E_1 \cos 30° = 2k \frac{|q|}{\varepsilon r^2} \frac{\sqrt{3}}{2} = k \frac{|q|\sqrt{3}}{\varepsilon r^2}.$$

Potential φ is equal to the sum of potentials $\varphi_1 = k \dfrac{q}{\varepsilon r}$ created by each of the charges:

$$\varphi = \varphi_1 + \varphi_1 = k \frac{2q}{\varepsilon r}.$$

2. A conducting sphere of radius $r = 0.2$ m is charged to a potential $\varphi = 3 \times 10^5$ V and placed in a dielectric with permittivity $\varepsilon = 27$. Determine: (1) the charge q of the sphere; (2) the magnitude E of the electric field strength on its surface; (3) the magnitude E_1 of the electric field strength and potential φ_1 at a point lying at a distance $r_1 = 10$ m from the centre of the sphere; (4) the magnitude E_0 of the field strength and potential φ_0 at the centre of the sphere.

Solution. The electric field of the charged sphere outside it coincides with the field of a point charge. Hence

$$\varphi = k \frac{q}{\varepsilon r}, \quad E = k \frac{|q|}{\varepsilon r^2}.$$

Consequently,

(1) $q = \dfrac{1}{k} \varepsilon r \varphi \simeq 1.8 \times 10^{-4}$ C;
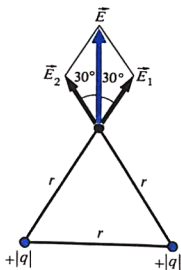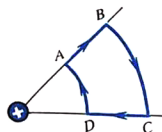

Fig. 131

Fig. 132

149

(2) $E = k \dfrac{|q|}{\varepsilon r^2} = \dfrac{\varphi}{r} \simeq 1.5 \times 10^6$ V/m;

(3) $E_1 = k \dfrac{|q|}{\varepsilon r_1^2} = \dfrac{r\varphi}{r_1^2}$ 600 V/m;

$\varphi_1 = k \dfrac{q}{\varepsilon r_1} = \dfrac{r\varphi}{r_1} \simeq 6000$ V.

(4) The charge $q > 0$ is distributed uniformly over the surface of the sphere. Inside the conducting sphere, the field strength at any point is zero: $E_0 = 0$. Therefore, the potentials of all points inside the sphere must be identical (and equal to the potentials of the points on the surface of the sphere): $\varphi_0 = \varphi$.

Exercise 9

1. A drop of liquid having a mass of 2×10^{-9} g is in equilibrium in a vertical downward electric field of strength 1.3×10^5 V m. Find the charge of the drop and the number of excess electrons on it.
2. Plot the magnitude of the electric field strength of a conducting charged sphere as a function of the distance from its centre.
3. Why does an electrostatically charged comb attract neutral pieces of paper?
4. An electric charge $q_1 > 0$ is moved over a closed path $ABCDA$ in the field of a point charge $q_2 > 0$ (Fig. 132). On what segments is the work of the field done in displacing the charge positive? negative? equal to zero? What is the change in the potential energy of the system? What is the total work done in displacing the charge?
5. An electron moving in an electric field passes from one point to another point whose potential is 1 V higher. What is the change in the kinetic and potential energies of the electron?
6. Point charges $q_1 > 0$ and $q_2 > 0$ are located at two vertices of an equilateral triangle with side r. Find the magnitude of the electric field strength at the third vertex if the permittivity of the medium is ε.
7. The potential of an electrostatic field increases in the upward direction. What is the direction of the electric field vector?
8. The potential difference between two points lying on the same line of force at a distance of 3 cm is 120 V. Find the electrostatic field strength if the field is known to be uniform.
9. Construct equipotential surfaces of an infinite conducting uniformly charged cylinder.
10. An electron moving in an electric field has increased its velocity from $v_1 = 1 \times 10^7$ m/s to $v_2 = 3 \times 10^7$ m/s. Determine the potential difference between the initial and final points of the electron path. The ratio of the electron charge to its mass is $e/m = 1.76 \times 10^{11}$ C/kg.

8.19. Capacitance. Units of Capacitance

Irrespective of the way in which a charge is imparted to a body by friction, with the Wimshurst machine or a voltaic cell, the initially neutral bodies are charged due to the transfer of a certain number of charged particles from one body to another. Usually these particles are electrons.

Let us clarify the question important for practical applications: under what condition can a large electric charge be accumulated on conducting bodies?

When two conductors are electrostatically charged (for example, from the Wimshurst machine), one of them acquires a charge $+|q|$ and the other $-|q|$. An electric field emerges between the conductors, and a potential difference (voltage) appears. With increasing voltage, the electric field between the conductors becomes stronger.

In a strong electric field (at a high voltage), a dielectric (e. g., air) becomes a conductor. The so-called "breakdown" of the dielectric is observed: a spark is struck between the conductors (Fig. 133) and they are discharged. The slower the rate at which the voltage between the conductors grows, the larger charge can be accumulated on them.

CAPACITANCE. Let us introduce a physical quantity characterizing the ability of two conductors to accumulate electric charge. This quantity is known as ELECTRIC CAPACITANCE.

The voltage U between two conductors is proportional to the electric charge on the conductors ($+|q|$ *on one of them and* $-|q|$ *on the other*). This follows from the principle of superposition. If the charges are doubled, the electric field strength becomes twice larger. Consequently, the work done by the field in displacing a charge (and hence the voltage) also increases twice. Therefore, the ratio of charge q of one of the conductors (the magnitude of the charge on the other conductor is the same) to the potential difference between the two conductors does not depend on the charge. It is determined only by geometrical sizes of the conductors, their shape and mutual arrangement as
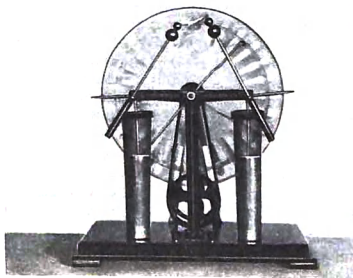

Fig. 133

well as by the electric properties of the surrounding medium (permittivity ε). This allows us to introduce the concept of electric capacitance of two conductors.

The capacitance of two conductors is the ratio of the charge of one of them to the potential difference between these conductors:

$$C = \frac{q}{U}.$$

(8.30)

The lower the voltage U created as a result of imparting charges $+|q|$ and $-|q|$ to the conductors, the larger their capacitance. Large charges can be accumulated on conductors without causing a dielectric breakdown. The capacitance itself, however, does not depend on the charges imparted to the conductors or on the voltage emerging as a result.

Sometimes we speak about the capacitance of a single conductor. This makes sense if the conductor is isolated, i.e. located at large distances from other conductors. For example, we may speak of the capacitance of a conducting sphere. It is assumed in this case that the role of the other conductor is played by remote objects around the sphere.

UNITS OF CAPACITANCE. Formula (8.30) allows us to introduce the unit of capacitance.

The electric capacitance of two conductors is equal to unity if the potential difference of 1 V appears between them when charges of $+1$ C and -1 C are imparted to them. This unit is called a farad (F). 1 F = 1 C/V.

Since the charge of 1 C is very large, the capacitance of 1 F is very large too. Therefore, the fractions of this unit are used in practice: microfarad (μF) equal to 10^{-6} F and picofarad (pF) equal to 10^{-12} F.

8.20. Capacitors. Capacitance of a Parallel-plate Capacitor

Systems of two conductors, called capacitors, have a high capacitance. A capacitor is a system of two conductors separated by a dielectric layer whose thickness is small in comparison with the size of the conductors. The conductors are called the CAPACITOR PLATES. A simple parallel-plate capacitor consists of two identical parallel plates separated by a small distance (Fig. 134). If the charges of the plates are equal and unlike, the lines of force of the electric field originate on the positively charged capacitor plate and terminate on the negative plate. Therefore, almost the entire electric field is concentrated inside the capacitor.

The electric field of a spherical capacitor consisting of two concentric spheres is concentrated completely between the spheres.

In order to charge a capacitor, its plates must be connected to the poles of a source of voltage, e.g. to the poles of a battery of accumulators. We can also connect one plate to a pole of a battery with the other pole being earthed, and earth the other plate. Then the charge remaining on the earthed plate will be equal and unlike to the charge of the other plate. The charge of the same magnitude passes to the earth.

The charge of a capacitor is assumed to be equal in magnitude to the charge on one of its plates.

The capacitance of a capacitor is defined by formula (8.30).

The electric fields of surrounding bodies practically do not penetrate inside a capacitor and do not affect the potential difference between its plates. Therefore, the capacitance of the capacitor virtually does not depend on the presence of other bodies in its neighbourhood.

The first capacitor, called the Leyden jar, was constructed in the middle of the 18th century. It was found that a nail immersed into a glass jar containing mercury stores a large electric charge. In such a capacitor, mercury served as a plate and the palms of an experimenter holding the jar as the other plate. Later, the two plates were manufactured of thin brass or tin foils.

CAPACITANCE OF A PARALLEL-PLATE CAPACITOR. Let us calculate the capacitance of a parallel-plate capacitor. We denote the surface area of each plate by S and the separation between the plates by d. To use formula (8.30) for calculating the capacitance, we must express the potential difference U in terms of charge q. This potential difference is determined by the strength E of the field, which depends on the charge of capacitor plates.

The field strength E_1 due to one of the plates is calculated by formula (8.16). The field strengths of the positive and negative plates are equal in magnitude and have the same direction inside the capacitor. Therefore, the magnitude E of the resultant field strength is equal to the sum of the magnitudes of the field strengths of the two plates:

$$E = 2E_1 = k\frac{4\pi|\sigma|}{\varepsilon}.$$

The formula for the capacitance can be written in a more compact form if we use its expression (8.6) instead of the coefficient k: $k = 1/(4\pi\varepsilon_0)$. Then, considering that the surface charge density $\sigma = q/S$, we obtain
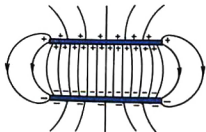
$$E = \frac{q}{\varepsilon_0\varepsilon S}.$$

Fig. 134
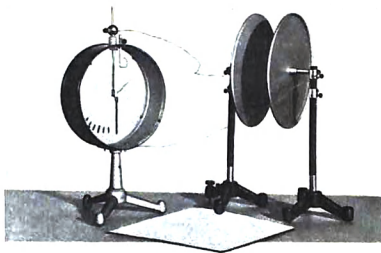

Fig. 135

Consequently, $U = Ed = qd/(\varepsilon_0 \varepsilon S)$.

Substituting this expression into (8.30) and cancelling out q, we obtain the formula for the capacitance of a parallel-plate capacitor:

$$C = \frac{\varepsilon_0 \varepsilon S}{d}. \qquad (8.31)$$

This formula shows that the capacitance of the capacitor depends on the geometrical factors, viz. the area of the plates and their separation, as well as on the electric properties of the medium. It does not depend on the material of the conductors: the capacitor plates can be made of iron, copper, aluminium, etc.

Let us verify experimentally the theoretically obtained formula (8.31). For this we take a capacitor whose plate separation can be varied and an electrometer with the earthed shell (Fig. 135). We connect the shell and rod of the electrometer with the capacitor plates and charge the capacitor. For this we can touch the capacitor plate connected to the rod with an electrostatically charged rod. The electrometer will indicate the potential difference between the plates.

Moving the plates apart, we see that the potential difference increases. According to definition (8.30) of capacitance, this points to its decrease. By formula (8.31), the capacitance must indeed decrease with increasing separation of the plates.

Placing a plate made of a dielectric, say, organic glass, between the capacitor plates, we shall observe a decrease in the potential difference. Consequently, the capacitance of the capacitor must increase.

The separation d of the plates can be very small, while the area S and permittivity ε quite large. Therefore, a capacitor may have a large capacitance even if its size is small. However, a parallel-plate capacitor having a capacitance of 1 F should have the area of plates $S = 100$ km^2 at the plate separation of 1 mm.

MEASUREMENT OF PERMITTIVITY. The dependence of the capacitance of a capacitor on the electric properties of a substance between its plates is used for measuring permittivity of substances. For this it is necessary to determine experimentally the ratio of the capacitance of the
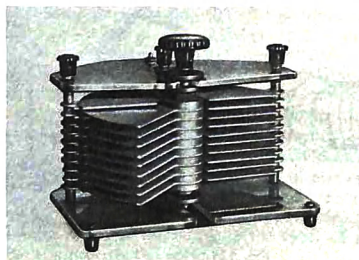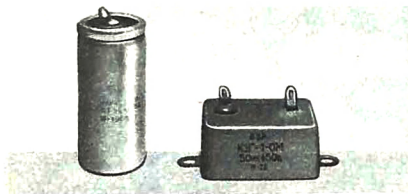

Fig. 136

Fig. 137

capacitor with the dielectric plate between the plates (C) and without it (C_0). It follows from formula (8.31) that the permittivity is

$$\varepsilon = \frac{C}{C_0}.$$

TYPES OF CAPACITORS. Capacitors have different construction depending on their duty. An ordinary paper capacitor used in radio engineering consists of two strips of aluminium foil, insulated from each other and from the metallic shell by paper bands impregnated by paraffin. The strips and bands are coiled into a small-size tube.

In radio engineering capacitors of variable capacitance (Fig. 136) are widely used. Such a capacitor consists of two systems of metallic plates. One system can be inserted in the gaps between the plates of the other system by rotating a handle. Thus the area of overlapping parts of the plates can be varied, and hence the capacitance too. In this type of capacitors, air is used as a dielectric.

In so-called electrolytic capacitors (Fig. 137), the capacitance can be considerably increased as a result of decreasing the separation between the plates. In this type of capacitors the dielectric is a thin oxide film coating one of the plates (foil strip). The second plate is paper impregnated by the solution of a special substance (electrolyte).

8.21. Energy of Charged Capacitor. Application of Capacitors

THE ENERGY OF A CHARGED CAPACITOR. In order to charge a capacitor, the work on separation of positive and negative charges should be done. According to the energy conservation law, this work is equal to the energy acquired by the capacitor.

The fact that a charged capacitor has an energy like any other system of charged bodies can be verified by discharging the capacitor through a circuit containing an incandescent lamp designed for a voltage of several volts (Fig. 138). When the capacitor discharges, the lamp glows brightly. The energy of the capacitor is transformed into other forms: thermal and light energy.
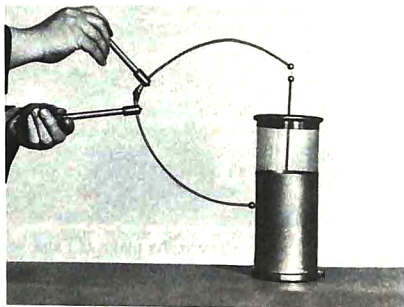
155

Fig. 138

Let us derive the formula for the energy of a parallel-plate capacitor. Suppose that the strength of the field produced by the charge on one plate is $E/2$, where E is the field strength in the capacitor (see Sec. 8.20). The charge q distributed over the surface of one plate is in the uniform field of the other plate (Fig. 139). According to formula (8.20) for the potential energy of a charge in a uniform field, the energy of the capacitor is

$$W_\text{p} = q \frac{E}{2} d, \tag{8.32}$$

where q is the capacitor charge and d is the separation of the plates[1].

Since $Ed = U$ is the potential difference between the capacitor plates, the energy of the capacitor is given by

$$W_\text{p} = \frac{qU}{2}. \tag{8.33}$$

This energy is equal to the work done by the electric field in moving the plates towards each other till they come in contact.

Replacing in formula (8.33) either the potential difference or charge with the help of expression (8.30) for the capacitance of a capacitor, we obtain

$$W_\text{p} = \frac{qU}{2} = \frac{q^2}{2C} = \frac{CU^2}{2}. \tag{8.34}$$

We can prove that these formulas are valid for the energy of any capacitor and not only a parallel-plate one.

---

[1] Formula (8.20) is valid for the energy of a point charge in a uniform field. However, we can mentally divide the charge on a plate into small elements Δq. The energy of each element is $\Delta W_\text{p} = \Delta q (E/2) d$. Summing up these energies, we arrive at formula (8.32).
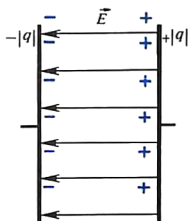
Fig. 139

THE ENERGY OF ELECTRIC FIELD. In accordance with the short-range theory, the entire energy of interaction between charged bodies is concentrated in the electric field of these bodies. Hence, the energy can be expressed in terms of the basic characteristic of the field, viz. its strength.

Let us substitute the expression for the capacitance of a parallel-plate capacitor [see (8.31)] into formula (8.34) and express the potential difference in this formula in terms of the field strength: $U = Ed$. Then the energy of the capacitor has the form

$$W_{\mathrm{p}} = \frac{CU^2}{2} = \frac{\varepsilon_0 \varepsilon S}{d} \cdot \frac{E^2 d^2}{2} = \frac{\varepsilon_0 \varepsilon E^2}{2} Sd. \tag{8.35}$$

Dividing (8.35) by the volume Sd occupied by the field, we obtain the energy per unit volume, i.e. energy density:

$$w_{\mathrm{p}} = \frac{\varepsilon_0 \varepsilon E^2}{2}. \tag{8.36}$$

This formula is valid not only for the uniform field of a parallel-plate capacitor, but for any other electrostatic field as well. Moreover, the obtained expression for the energy density proves to be valid for varying electric fields too.

APPLICATION OF CAPACITORS. The energy of a capacitor is normally not very high (not higher than hundreds of joules). Besides, it cannot be stored for a long time because of charge leakage. For this reason, charged capacitors cannot replace, for example, accumulators as sources of electric power.

This does not mean, however, that capacitors have not found any practical application as accumulators of energy. They possess a very important property. A capacitor may accumulate energy for quite a long time, but during a discharge through a low-resistance circuit, it releases energy almost instantaneously. This property is widely used for practical purposes.

A flash-lamp used in photography operates on the discharge current of a capacitor which is preliminary charged by a special battery. The excitation of quantum sources of light, viz. lasers, is realized with the help of gas-discharge tubes which burst as a result of the discharge of a capacitor battery of a high capacitance.

But the broadest field of application of capacitors is radio engineering.

Capacitors are used in various electric curcuits to obtain varying voltage by changing a charge. High-capacitance capacitors are able to accumulate or release large charges without a noticeable change in voltage.

1. Under which condition can a large electric charge be accumulated on a conductor?
2. What is called the capacitance of two conductors?
3. Why is the concept of capacitance inapplicable to dielectrics?
4. What does capacitance depend on?
5. Give the definition of the unit of capacitance.
6. Write the expression for the capacitance of parallel-plate capacitor
7. What is the energy of a capacitor equal to?
8. List the main applications of capacitors.

Problems with Solutions

Capacitance is the last topic in electrostatics. While solving problems, you may require the concepts and laws of this branch of electrostatics, viz. the law of electric charge conservation, the concepts of field strength and potential, the behaviour of conductors in an electrostatic field, the effect of dielectrics on the field, and the energy conservation law for electrostatic phenomena. The solution of problems involving capacitance will not present any difficulty if you have a clear idea about all basic concepts of electrostatics.

The main formulas required for solving problems on capacitance are the definition (8.30) of capacitance and the expression (8.31) for capacitance of a parallel-plate capacitor.

1. The capacitance of a capacitor is $C = 5$ pF. Determine the charge on each of its plates if the potential difference across them is $U = 1000$ V.

Solution. According to formula (8.30), the capacitance of a capacitor is $C = q/U$. Hence the charge of a plate is $q = CU$,

$$q = 5 \times 10^{-12} \times 1000 \text{ C} = 5 \times 10^{-9} \text{ C}.$$

2. Determine the capacitance of the parallel-connected capacitors shown in Fig. 140. (In a parallel connection positively and negatively charged plates of the capacitors are connected pairwise.)

Solution. The potential difference across the capacitor plates is the same for both capacitors. If the charge of the first capacitor is q_1 and that of the second is q_2, then their total charge is $q = q_1 + q_2$ and their capacitance is $C = q/U = (q_1 + q_2)/U$. Since $C_1 = q_1/U$ and $C_2 = q_2/U$, the total capacitance of two parallel-connected capacitors is

$$C = C_1 + C_2. \tag{8.37}$$

3. Two capacitors with capacitances C_1 and C_2 are series-connected (Fig. 141). Determine the capacitance of the capacitor bank. (In series connection the negatively charged plate of the first capacitor is connected to the plate of the second capacitor, carrying a positive charge of the same magnitude.)
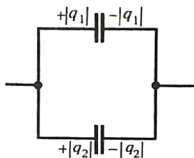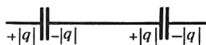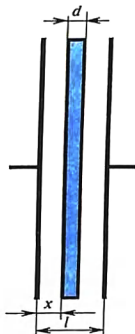
Fig. 140

Fig. 141

Fig. 142

Solution. In the case of a series connection of capacitors their charges are equal. Indeed, if the charge of the first plate of the first capacitor is $+|q|$, the charge $-|q|$ appears on its opposite plate. Since the wire between the capacitors and the plates connected by it are on the whole electrically neutral, the inner plate of the second capacitor acquires the charge $+|q|$.

The total capacitance of series-connected capacitors is $C = q/(\varphi_1 - \varphi_2)$, where φ_1 and φ_2 are the potentials of the outer plates. The capacitances of the capacitors are $C_1 = q/(\varphi_1 - \varphi')$ and $C_2 = q/(\varphi' - \varphi_2)$, where φ' are the potentials of the inner plates of the capacitors. We find the reciprocal values of the capacitances:

$$\frac{1}{C_1} + \frac{1}{C_2} = \frac{\varphi_1 - \varphi'}{q} + \frac{\varphi' - \varphi_2}{q} = \frac{\varphi_1 - \varphi_2}{q} = \frac{1}{C}.$$

Hence, in a series connection of capacitors the reciprocal value of the capacitance of a capacitor bank is equal to the sum of the reciprocal values of capacitances of separate capacitors

$$\frac{1}{C} = \frac{1}{C_1} + \frac{1}{C_2}. \tag{8.38}$$

Exercise 10

1. Determine the capacitance of a parallel-plate capacitor whose plates are two foil strips having an area $S = 0.4$ m^2 each and separated by paraffin paper ($\varepsilon = 2.2$) having a thickness $d = 0.08$ mm.
2. The potential difference across the capacitor plates described in the preceding problem has changed by 175 V. Determine the change in the capacitor charge.
3. The charge of a parallel-plate capacitor with a mica dielectric is 2.7×10^{-4} C. The area of each plate is 2500 cm^2. The dielectric

159

permittivity of mica is 7. Calculate the field strength in the dielectric.

4. An electron having a velocity of 2×10^7 m/s flies into the space between the plates of a parallel-plate capacitor so that its velocity vector is parallel to the plates. By what distance is the electron displaced towards the positive plate during its motion within the capacitor if its capacitance $C = 0.885$ pF, the length of the capacitor plates is 0.05 m, and the potential difference across the plates is 200 V? The surface area of the capacitor plates is $S = 0.002$ m^2. The charge-to-mass ratio for the electron is 1.76×10^{11} C/kg.

5. A parallel-plate capacitor is charged with the help of a source having a voltage $U = 200$ V and then disconnected from the source. What will be the voltage U_1 across the plates if their separation is increased from $d = 0.2$ mm to $d_1 = 0.7$ mm, and the space between the plates is filled with mica. The permittivity of mica, is $\varepsilon = 7$.

6. Find the capacitance C of a capacitor if the area of its plates is S, the distance between them is l, and a metallic plate of thickness d and parallel to the plates of the capacitor is inserted into it (Fig. 142).

7. What will be the change in the energy of a charged capacitor disconnected from a source if the space between its plates is filled with a dielectric having a permittivity $\varepsilon = 3$?

8. Find the energy density of the electric field of the capacitor described in Problem 4.

Main Points of Chapter 8

According to Coulomb's law fixed point charges q_1 and q_2 interact in vacuum with the force

$$F = k \frac{|q_1||q_2|}{r^2},$$

where the coefficient $k = 9 \times 10^9$ N·m^2/C^2. The charges are measured in coulombs.

In a closed system of charged particles, the electric charge is conserved.

Charges interact via the electric field. The field strength $\vec{E}$ determines the force acting on a charge: $\vec{F} = q\vec{E}$. The strengths of the fields produced by individual charges are added geometrically (principle of superposition).

The electric field strength of a point charge in vacuum is given by

$$E = k \frac{|q|}{r^2}.$$

The field strength for a plane is $E = k\, 2\pi |\sigma|$, where σ is the surface charge density.

Conductors contain free electric charges. The field strength and the charge inside a conductor are zero (in electrostatics).

In dielectrics, charges are bound inside individual atoms or molecules. A homogeneous dielectric weakens the field by a factor of ε, where ε is the dielectric permittivity.

The electrostatic field is a potential field: its work does not depend on the shape of the path of a charge and is equal to the change in the potential energy with the minus sign: $A = -\Delta W_p$.

The potential energy of a charge in a uniform field is $W_p = qEd$, where d is the distance from the plane on which the potential energy is assumed to be zero.

The electric field potential is the ratio of the potential energy of a charge in the field to this charge:

$$\varphi = \frac{W_p}{q}.$$

The potential difference (voltage) between two points is equal to the ratio of the work done in displacing a charge from the initial to the final point to this charge: $U = \varphi_1 - \varphi_2 = A/q$. The potential difference is expressed in volts: $1\ V = 1\ J/1\ C$. Potential of a point charge in vacuum is

$$\varphi = k\frac{q}{r}.$$

The field strength is connected with the potential difference through the following relation:

$$E = \frac{U}{\Delta d},$$

where U is the potential difference between two points lying on the same line of force at a small distance Δd from each other.

The ability of conductors to store electric charge is characterized by electric capacitance. The capacitance of two conductors is given by

$$C = \frac{q}{U},$$

where q is the charge of one of the conductors (the other conductor bears the charge of the opposite sign), and U is the potential difference between the conductors.

Capacitance is expressed in farads:

$$1\ F = 1\ C/1\ V.$$

Large charges can be stored in capacitors, viz. systems of two conductors whose size is much larger than their separation.

The capacitance of a parallel-plate capacitor is given by

$$C = \frac{\varepsilon_0 \varepsilon S}{d},$$

where S is the area of the plates, d is their separation, and $\varepsilon_0 = 1/4\pi k$ is the electric constant.

The energy of a charged capacitor is $W_p = qU/2$. The electric field energy density is given by

$$w_p = \frac{\varepsilon_0 \varepsilon E^2}{2}.$$

We devoted a considerable space of the book to consider only the simplest special case of fixed charged bodies, viz. electrostatics. It may appear that it does not deserve so much attention. But it does! We have introduced the basic concepts used throughout electrodynamics: electric charge, electric field, potential and potential difference, capacitance and energy of electric field. It is easier to elucidate the meaning of these fundamental concepts in a simple particular case than in the general case of moving charges.

We now go over to studying electromagnetic processes observed during the motion of charged particles.

9 DIRECT CURRENT

9.1. Electric Current

During the motion of charged particles electric charge is transferred from one site to another. If, however, charged particles are in random thermal motion like, for example, free electrons in a metal, no charge transfer is observed (Fig. 143). An electric charge is transferred through the cross section of a conductor only if electrons take part in ordered motion in addition to chaotic movement (Fig. 144). In this case, ELECTRIC CURRENT is said to set in in the conductor.

It was mentioned in the course of "Junior Physics" that electric current is the ordered (directional) motion of charged particles. emerges during the displacement of free electrons in a metal or ions in electrolytes.

If, however, a neutral (on the whole) body is displaced, no electric current appears in spite of the ordered motion of a huge number of electrons and atomic nuclei. The total charge transferred through any cross section of the conductor is zero in this case since unlike charges are moved with the same average velocity. A current appears in a conductor only if the positive charge transported through the conductor's cross section during the motion of charges in one direction is not equal in magnitude to the negative charge.

Electric current has a certain direction. The direction of current is defined as the direction of motion of positively charged particles. *If a current is formed by the motion of negatively charged particles, it is assumed that its direction is opposite to the direction of motion of particles.*[1]

EFFECTS OF CURRENT. We cannot observe directly the motion of particles in a conductor. However, the presence of electric current is manifested in effects and phenomena accompanying the current.

First, *a current-carrying conductor is heated.*

---

[1] Such a choice of the direction of current is not very convenient since in most cases the current is the motion of electrons, viz. negatively charged particles. This direction of current was chosen when the existence of free electrons in metals was unknown.
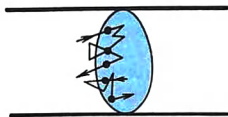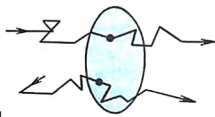
Fig. 143   Fig. 144

Second, *the electric current can change the chemical composition of a conductor*, for example, to separate its chemical components (copper from a copper sulphate solution, and so on). The processes of this kind are observed only in certain conductors like solutions (or melts) of electrolytes.[1]

Third, *the current exerts a force on neighbouring currents and magnetized bodies*. This effect is known as the magnetic effect. For example, a magnetic needle turns in the vicinity of a current-carrying conductor. Unlike the chemical and thermal effects of current, the magnetic effect is of principal nature since it is observed for all conductors without exception. The chemical action of current is observed only for solutions or melts of electrolytes, while heating does not take place in superconductors (Sec. 9.5).

CURRENT. If an electric current is sustained in a circuit, an electric charge is continuously transferred through the cross section of the conductors. The charge transported per unit time is the main quantitative characteristic of the current and is known as CURRENT. If a charge Δq is transferred through the cross section of a conductor in time Δt, the current is

$$I = \frac{\Delta q}{\Delta t}. \qquad (9.1)$$

Thus, the current is equal to the ratio of the charge Δq transferred through the conductor cross section during the time interval Δt to this time. If a current does not change with time, it is called CONSTANT, or DIRECT current.

Like charge, current is a scalar quantity. It can be either positive or negative. The sign of the current depends on the direction along the conductor taken as positive. The current $I > 0$ if the direction of current coincides with the conditionally chosen positive direction along the conductor. Otherwise, $I < 0$.

The current depends on the charge transported by each particle, the number density of particles, the velocity of their directional motion, and the cross-sectional area of the conductor. Let us prove this.

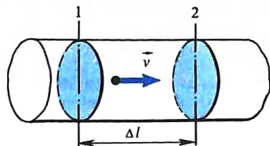[1] For details, the textbooks on "Inorganic Chemistry".

Fig. 145

Suppose that the cross-sectional area of a conductor is S. For the positive direction, we choose the direction from left to right. The charge of each particle is q_0. The number of particles contained in the volume of the conductor bounded by sections 1 and 2 is $nS\,\Delta l$, where n is the number density of particles (Fig. 145). The total charge of these particles is $q = q_0 nS\,\Delta l$. If the particles move from left to right at an average velocity $v^{1)}$, all particles contained in this volume will pass through section 2 in time $\Delta t = \Delta l/v$. Therefore, the current is

$$I = \frac{q}{\Delta t} = \frac{q_0 nvS\,\Delta l}{\Delta l} = q_0 nvS. \qquad (9.2)$$

In SI, current is expressed in amperes (A). This unit is established from the magnetic interaction of currents. Current is measured by the instruments called AMMETERS. The operational principle of these instruments is based on the magnetic effect of current and will be described later.

THE VELOCITY OF ORDERED MOTION OF ELECTRONS IN A CONDUCTOR. Let us determine the velocity of ordered displacement of electrons in a metallic conductor.

According to formula (9.2), we have

$$|v| = \frac{|I|}{enS},$$

where e is the magnitude of the electron charge. Let, for example, the current I be equal to 1 A, and the cross-sectional area of the conductor be $S = 10^{-6}$ m^2. The magnitude of the electron charge is $e = 1.6 \times 10^{-19}$ C. The number of electrons in one cubic metre of copper is equal to the number of atoms in this volume since one of the valence electrons in each copper atom is collectivized and becomes free. This number is $n = 8.5 \times 10^{28}$ m^{-3} (see Problem 7 in Ex. 1). Consequently,

$$|v| = \frac{1}{1.6 \times 10^{-19} \times 8.5 \times 10^{28} \times 10^{-6}} \text{ m/s} = 7 \times 10^{-5} \text{ m/s}.$$

The velocity of ordered motion of electrons appears to be very small.

9.2. Conditions Required for the Existence of Electric Current

To generate and maintain a direct electric current in a substance, the presence of free charged particles is required above all. If positive and negative charges are bound in atoms or molecules, their displacement would not lead to an electric current.

However, the presence of free charges is insufficient for the current to appear. To create and sustain an ordered motion of charged particles, a force acting on them in a certain direction is required. If this force ceases to act, the ordered motion of charged particles will stop due to the resistance

---
[1] To be more precise, v is the projection of the average velocity onto the positive direction along the conductor.
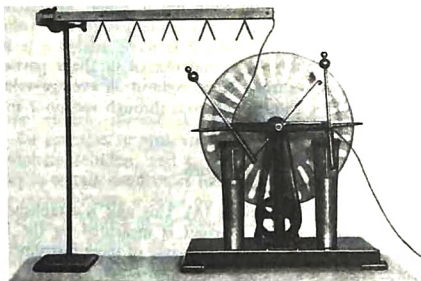
Fig. 146

offered to the motion by the ions constituting the crystal lattices of metals or by neutral molecules of electrolytes.

As is well known, an electric field acts on charged particles with the force $\vec{F} = q\vec{E}$. *Normally, the electric field in the conductor is just the cause of creating and sustaining the ordered motion of charged particles.* The electric field is zero inside a conductor only in the static case when the charges are at rest.

IF THERE EXISTS AN ELECTRIC FIELD INSIDE A CONDUCTOR, THEN, ACCORDING TO FORMULA (8.29), THERE EXISTS A POTENTIAL DIFFERENCE ACROSS THE ENDS OF THE CONDUCTOR. If this potential difference does not vary with time, a constant electric current is sustained in the conductor. Along the conductor, the potential drops from the maximum value at one end of the conductor to the minimum value at the other end. This decrease in potential can be demonstrated by a simple experiment.

Let us take a conductor in the form of a not very dry wooden stick and suspend it in the horizontal position. (Such a stick conducts current, although it is a poor conductor.) We shall use the Wimshurst machine as a source of voltage. The potential of the segments of the stick relative to the Earth can be registered with the help of the strips of tin foil fixed to the stick. We earth one pole of the machine and connect the other pole to the end of the conductor (stick). The circuit is open. Rotating the handle of the machine, we see that all the strips diverge by the same angle (Fig. 146). This means that the potential of all points of the conductor relative to the earth is the same. This is what should be expected for charges in equilibrium. If we now earth the other end of the stick and rotate the handle of the machine again, the situation will change. (The earthing of the conductor makes the circuit closed since the Earth is a conductor.) The strips at the earthed end do not diverge at all, since the potential of this point is virtually equal to the potential of the Earth (the voltage drop across the connecting wire is very small). The maximum angle of divergence will be at the point of the conductor connected to the machine (Fig. 147). The decrease in the divergence angle with increasing distance to the machine indicates a potential drop along the conductor.

166

Fig. 147
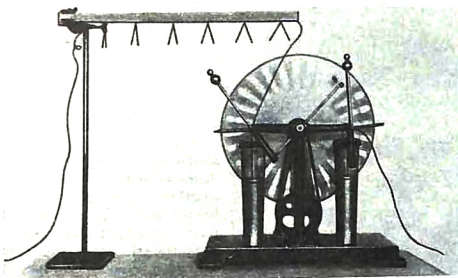
1. What is called the electric current?
2. Which direction of current is taken as positive?
3. What conditions are required for the existence of electric current?

9.3. Ohm's Law for a Conductor. Resistance

VOLTAGE-CURRENT CHARACTERISTIC. It was established in the preceding section that to sustain a current in a conductor, it is necessary to create a potential difference across its ends. The current in the conductor is determined by this potential difference. The larger the potential difference, the higher the electric field strength in the conductor, and hence the higher the velocity of directional motion acquired by charged particles. According to (9.2), this points to an increase in the current.

For each conductor (solid, liquid, or gaseous) there exists a certain dependence of the current on the applied potential difference. This dependence is expressed by the so-called VOLTAGE-CURRENT CHARACTERISTIC OF THE CONDUCTOR. It is obtained by measuring the current in the conductor at different values of the voltage. The knowledge of the voltage-current characteristic is very important for the analysis of phenomena involving the passage of current.

OHM'S LAW. The voltage-current characteristic for metallic conductors and electrolyte solutions has a simple form. This dependence (for metals) was established for the first time by the German scientist G. OHM. For this reason, the dependence of current on voltage is known as Ohm's law. This law was described in detail in the course of "Junior Physics", but it is so important that it deserves a more detailed discussion.

In the subcircuit shown in Fig. 148, the current is directed from point 1 to point 2. The potential difference (voltage) across the ends of the conductor is $U = \varphi_1 - \varphi_2$. Since the current is directed from left to right, the electric field strength has the same direction, and $\varphi_1 > \varphi_2$.

167

According to Ohm's law, the current in a conductor is directly proportional to the applied voltage U and inversely proportional to the resistance R of the conductor:

$$I = \frac{U}{R}. \tag{9.3}$$

Although this law has a very simple form, its experimental verification presents considerable difficulties. As a matter of fact, even at large currents the potential difference across a segment of a metallic conductor is small as well as the resistance of the conductor. The electrometer described in Sec. 8.18 is inapplicable for measuring such small voltages since its sensitivity is too low. A much more sensitive instrument is required for this purpose. By measuring the current with an ammeter and voltage with a sensitive electrometer, we can show that the current is proportional to the voltage. The operation of conventional instruments for measuring voltage, viz. voltmeters, is based on Ohm's law, and hence such instruments cannot be used here.

The working principle of a voltmeter is the same as that of an ammeter. The angle of deflection of the pointer is proportional to the current. The current passing through the voltmeter is determined by the voltage between the points of the circuit to which it is connected. Therefore, if we know the resistance of the voltmeter, we can determine the voltage from the current. In
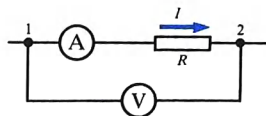

Fig. 148

practice, the instrument is graduated so that it directly indicates voltage in volts.

RESISTANCE. Resistance is the basic characteristic of a conductor. This quantity determines the current in the conductor at a given voltage. The resistance of a conductor is a measure of counteraction of the conductor to the current through it. The resistance can be determined with the help of Ohm's law (9.3): $R = U/I$. For this purpose, we must measure the voltage and the current.

The resistance depends on the material of which the conductor is made and on its geometrical dimensions. The resistance of a conductor having a length l and the cross-sectional area S is

$$R = \rho \frac{l}{S},$$

where ρ is a quantity depending on the kind of the substance and its state (above all, the temperature). This quantity is known as the RESISTIVITY of the conductor. The resistivity is numerically equal to the resistance of a conductor of cubic shape with an edge of 1 m, if the current is directed along the normal to two opposite faces of the cube.

The unit of resistance is established on the basis of Ohm's law and is called the ohm. A conductor has a resistance of 1 Ω if the current in it is 1 A for a potential difference of 1 V.

The unit of resistivity is 1 Ω·m. The resistivity of metals is very low. Dielectrics have a very high resistivity. Resistivities of several materials are compiled in tabular form on the end sheet of this book.

THE IMPORTANCE OF OHM'S LAW. Ohm's law determines the current in an electric circuit for a given voltage and known resistance. Thus, it allows us to calculate thermal, chemical and magnetic effects since they are determined by the current. It follows from Ohm's law that it is dangerous to close an ordinary lighting system by a small-resistance conductor. According to (9.3), the current will then be so large that this may lead to serious consequences.

9.4. Temperature Dependence of the Resistance of a Conductor

The resistance of a conductor varies with temperature. This can be verified by passing a current from an accumulator through a steel spiral heated in the flame of a burner. An ammeter in the circuit will indicate a decrease in the current with increasing temperature.

If the resistance of a conductor is R_0 at 0 °C and R at a temperature t, experiments show that the relative change in resistance is proportional to the change in temperature t:

$$\frac{R - R_0}{R_0} = \alpha t. \tag{9.4}$$

The proportionality factor α is known as the TEMPERATURE COEFFICIENT OF RESISTANCE. It characterizes the temperature dependence of the resistance

of a material. *The temperature coefficient of resistance is numerically equal to the relative change in the resistance of a conductor upon a change in its temperature. by* 1 K. For metallic conductors, $\alpha > 0$ and varies slightly with temperature. If the range of temperature is small, the temperature coefficient of resistance can be assumed to be constant and equal to the mean value for this temperature interval. For pure metals, $\alpha \simeq 1/273$ K^{-1}.

The resistance of electrolyte solutions decreases (and not increases) with temperature. For these conductors, $\alpha < 0$. For example, $\alpha = -0.02$ K^{-1} for a 10%-solution of common salt.

The geometrical dimensions of a conductor change insignificantly upon heating. Its resistance mainly changes due to a change in resistivity. We can find the temperature dependence of resistivity by substituting into formula (9.4) the values $R_0 = \rho_0 l/S$ and $R = \rho l/S$. As a result of calculations. we obtain

$$\rho = \rho_0(1 + \alpha t). \tag{9.5}$$

Since α changes with temperature insignificantly, we can assume that the resistivity linearly depends on the temperature (Fig. 149).

Although the coefficient α is quite small, it is necessary to take into account the temperature dependence of resistance in designing heating devices. For example, the resistance of the tungsten filament of a conventional lamp increases more than tenfold when a current is passed through it.

Some alloys, like the copper-nickel alloy Constantan, have a very small temperature coefficient of resistance: $\alpha \simeq 10^{-5}$ K^{-1}. The resistivity of Constantan is very high: $\rho \simeq 10^{-6}$ $\Omega \cdot m$. Such alloys are used for manufacturing standard resistances and series resistances for measuring instruments. i.e. when it is required that the resistance must not change noticeably with temperature variations.

The temperature dependence of resistance is used in RESISTANCE THERMOMETERS. The working element of such a thermometer is normally a platinum wire the temperature dependence of whose resistance is well known. The variation of temperature is judged from the change in the wire resistance which can be measured. Such thermometers make it possible to measure very low as well as very high temperatures for which ordinary thermometers are unsuitable.
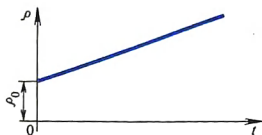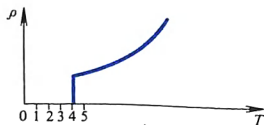

·Fig. 149

Fig. 150

9.5. Superconductivity

In 1911, the Dutch physicist H. Kamerling-Onnes discovered a wonderful phenomenon, viz. SUPERCONDUCTIVITY. He found that when mercury was cooled in liquid helium, its resistance first varies gradually, and then, at a temperature of 4.1 K, sharply drops to zero (Fig. 150). This phenomenon was called superconductivity. Later, many other superconductors were discovered.

Superconductivity is observed at very low temperatures not exceeding 25 K. The table on the end sheet gives the values of the temperatures of transition to the superconducting state for several substances.

If current is generated in a circular conductor in the superconducting state and then the current source is eliminated, a constant current continues to flow for an infinitely long time. On the other hand, in an ordinary nonsuperconducting conductor the current rapidly attenuates under similar conditions.

Superconductors have found a wide practical application. High-power electromagnets with superconducting windings are being constructed, which produce a magnetic field over long time intervals without energy expenditures, since no heat is liberated in the superconducting winding.

It is impossible, however, to obtain as strong magnetic field as desired with the help of a superconducting magnet. A very strong field destroys the superconducting state. Such a field may be produced by the current in the superconductor itself. Therefore for every conductor in the superconducting state there exists a critical value of the current which cannot be exceeded without violating the superconducting state.

Superconducting magnets are used in accelerators of elementary particles, and MHD (magnetohydrodynamic) generators which transform the mechanical energy of an incandescent ionized gas jet moving in a magnetic field into electric energy.

If superconductivity could be obtained at nearly room temperatures, an important engineering problem, viz. a loss-free energy transmission along wires, would have been solved. Physicists are working at this problem now.

Superconductivity can be explained only on the basis of the quantum theory. The explanation was proposed only in 1957 by the American scientists J. Bardeen, L. Cooper and J. Schrieffer and by the Soviet Academician N. N. Bogolyubov.

9.6. Electric Circuits. Series and Parallel Connection of Conductors

Energy can be transmitted from a current source through wires to devices like an electric bulb, electric motor, radio set, etc., which consume this energy. For this, electric circuits of various degrees of complexity are constructed. An electric circuit contains a power source, the devices consuming electric energy, connecting wires and switches for connecting the circuit. Electric circuits often contain measuring instruments

intended to control current and voltage in various subcircuits, viz. ammeters and voltmeters.

The simple and most frequently encountered types of connection of conductors are SERIES and PARALLEL CONNECTION.

SERIES CONNECTION OF CONDUCTORS. In the series connection, an electric circuit is not branched. All conductors are connected successively one after another. Figure 151 shows the series connection of two conductors *1* and *2* having resistances R_1 and R_2. These can be two bulbs, two windings of a motor, and so on.

The current is the same in both the conductors:

$$I_1 = I_2 = I, \tag{9.6}$$

since in the case of direct current the electric charge is not accumulated, and the same charge passes through any cross section of a conductor during a certain interval of time.

The voltage (or potential difference) across the ends of the subcircuit under consideration is the sum of the voltage across the first and second conductor: $U = U_1 + U_2$.

Applying Ohm's law for the subcircuit as a whole and for conductors having resistances R_1 and R_2, we can prove that the total resistance for the series connection is

$$R = R_1 + R_2. \tag{9.7}$$

A similar formula can be used for any number of series-connected conductors.

The voltages across the series-connected conductors and their resistances are connected through the following relation:

$$\frac{U_1}{U_2} = \frac{R_1}{R_2}. \tag{9.8}$$

PARALLEL CONNECTION OF CONDUCTORS. Figure 152 shows the parallel connection of two conductors *1* and *2* having resistances R_1 and R_2. In this case, the current I is branched into two parts. We denote by I_1 and I_2 the currents in the first and second conductors. Since no electric charge is accumulated at the branching point a (such a point is known as junction), the charge supplied per unit time to the junction is equal to the charge leaving the junction during the same time. Consequently,
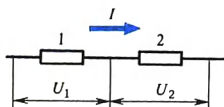
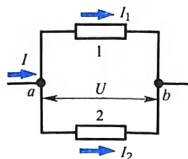$$I = I_1 + I_2. \tag{9.9}$$

Fig. 151

Fig. 152

The voltage U across the parallel-connected conductors is the same.

The voltage maintained in the lighting system in the USSR is either 220 or 127 V. The devices consuming the electric energy are designed for this voltage. For this reason, the parallel connection is the most widely used way of connecting various loads. In this case, a breakdown of any device does not affect the operation of others, while a breakdown of a single device in a series connection disconnects the entire circuit.

Applying Ohm's law for the conductors having resistances R_1 and R_2, we can prove that the reciprocal of the total resistance of subcircuit ab is equal to the sum of the reciprocals of the resistances of individual conductors:

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2}. \tag{9.10}$$

Hence it follows that

$$R = \frac{R_1 R_2}{R_1 + R_2}. \tag{9.11}$$

Formulas similar to (9.9) and (9.10) can be applied when an arbitrary number of conductors are connected in parallel.

There exists the following relation between the currents in parallel-connected conductors and their resistances:

$$\frac{I_1}{I_2} = \frac{R_2}{R_1}. \tag{9.12}$$

9.7. Measurement of Current and Voltage

MEASUREMENT OF CURRENT. The current in a conductor is measured by connecting an ammeter in series with this conductor (Fig. 153). It should be borne in mind, however, that the ammeter itself has a certain resistance R_a. Therefore, the resistance of the subcircuit containing the ammeter becomes higher, and according to Ohm's law (9.3), the current decreases if the voltage remains the same. To minimize the effect of the ammeter on the current which it measures, its resistance is made very small. This should be remembered, and one should never try to "measure current" in a lighting system by connecting an ammeter to the socket outlet. This would lead to short-circuiting, since resistance of the instrument is very low, the current would rise to such a high value that the ammeter winding would be burnt.

SHUNTS TO AMMETER. Every ammeter is designed for measuring currents up to a certain maximum value I_0 (e.g., 5 A). However, by connecting a resistor known as SHUNT in parallel to the ammeter, we can measure a stronger (for example, 10-fold) current: $I = nI_0$ ($I = 50$ A).

Let us calculate the resistance R_{sh} of the shunt required to measure a current n times stronger than the rated current. We denote the resistance of the ammeter by R_a.

When the shunt is used, a fraction I_{sh} of the current flows through it. The
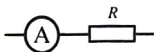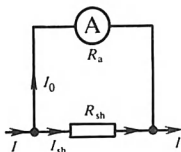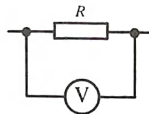
Fig. 153

Fig. 154

Fig. 155

current through the ammeter must not be higher than $I_0 = 5$ A (Fig. 154). The current I_0 is smaller than the current to be measured by a fraction of n: $I_0 = I/n$ (5 A = 50 A/10). Consequently, the current corresponding to one scale division becomes n (10) times larger, i.e. the deflection of the pointer by one division will correspond to an n times stronger current (10 A instead of 1 A). In other words, the sensitivity of the ammeter decreases by a factor of n: when the shunt is connected, the pointer is deflected by an angle smaller than that without shunt by a factor of n (10).

For the parallel connection, we have $I = I_0 n = I_0 + I_{sh}$, while the voltage is the same across the shunt and the ammeter and, according to Ohm's law, is $I_0 R_a = I_{sh} R_{sh}$. Eliminating current I_0 from the last two equations, we obtain

$$R_{sh} = \frac{R_a}{n - 1}. \tag{9.13}$$

MEASUREMENT OF VOLTAGE. In order to measure voltage across a conductor having a resistance R, we must connect a voltmeter in parallel to it. The voltage across the voltmeter coincides with the voltage across the conductor (Fig. 155).

If the resistance of the voltmeter is R_v, the resistance of the subcircuit after the voltmeter has been connected is not R but $R' = RR_v/(R + R_v) < R$. Hence the voltage to be measured across the subcircuit decreases. In order to minimize the distortions of the voltage being measured, the resistance of the voltmeter must be high in comparison with the resistance of the conductor across which the voltage is being measured. There is no risk in connecting a voltmeter to a circuit. It would not be burnt unless it is designed for a voltage exceeding that of the lighting system.

ADDITIONAL RESISTANCES IN SERIES TO A VOLTMETER. Any voltmeter is rated for a maximum resistance U_0 (say, 50 V). However, using a series-connected resistor whose resistance is R_s, the voltmeter can be used for measuring n (for example, 10) times higher voltages $U = nU_0$ (500 V = 10×50 V).

Let us determine the resistance of a series resistor required for measuring voltages n times higher than the rated values. We denote the resistance of the voltmeter by R_v.

When a series resistor is connected to the voltmeter subcircuit, the maximum resistance across the voltmeter must be, as before, U_0 (50 V), but it must be only $1/n$ (1/10) of the voltage being measured: $U_0 = (U/n)$ (50 V = 500 V/10). The remaining fraction of the voltage ($U_s = U - U_0 = 500$ V − 50 V = 450 V) falls on the series resistor (Fig. 156). Therefore, the limits of
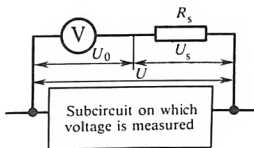
174

Fig. 156

measurement increase n (10) times, and the voltage corresponding to one scale division of the voltmeter increases accordingly (10 V instead of 1 V). Hence the sensitivity of the instrument becomes 10 times lower.

The same current flows through the voltmeter and the series resistor. Therefore, $U_0 = IR_v$, $U_s = IR_s$ and $U = U_0 + U_s = IR_v + IR_s = nU_0 = nIR_v$. Hence

$$R_s = R_v (n - 1). \qquad (9.14)$$

9.8. Work and Power of Direct Current

Many types of energy conversion take place in an electric circuit. The electric field does work in the ordered motion of charged particles in a conductor. This work is conventionally called the WORK OF CURRENT.

Let us consider an arbitrary subcircuit. This can be a homogeneous conductor like the filament of an incandescent lamp, the winding of a motor, etc. Suppose that a charge Δq passes through the cross section of the conductor in time Δt. The work done by the electric field is $A = \Delta qU$.

Since the current $I = \Delta q/\Delta t$, this work is

$$A = IU \Delta t. \qquad (9.15)$$

The work of current in a subcircuit is equal to the product of the current by the voltage and time during which this work is done.

In accordance with the energy conservation law, this work must be equal to the change in the energy in the subcircuit under consideration. Therefore, the energy liberated in this subcircuit over a time Δt is equal to the work of current (9.15).

If no mechanical work is done on the subcircuit and the current does not produce any chemical effects, only the heating of the conductor takes place. The heated conductor gives away heat to the surrounding bodies.

Heating of a conductor proceeds as follows. The electric field accelerates electrons. As a result of collisions with the ions constituting the crystal lattice, the electrons transfer their energy to the ions, thus intensifying the random movement of the ions about their equilibrium position. This means that the internal energy of the conductor increases. The temperature rises and the conductor starts to give away energy to the surrounding bodies. A short time after the circuit has been closed, the steady-state process is established, and the temperature stops changing with time. At the expense of work of the electric field, energy is continuously supplied to the conductor.

Its internal energy, however, remains unchanged since it gives away to the surroundings an amount of heat equal to the work of current. Thus, formula (9.15) for the work of current determines the amount of heat transferred by the conductor to other bodies.

If we express the voltage in formula (9.15) in terms of current (or current in terms of voltage) with the help of Ohm's law, we obtain three equivalent formulas for the work of current:

$$A = IU \Delta t = I^2 R \Delta t = \frac{U^2}{R} \Delta t = Q. \tag{9.16}$$

It is convenient to use the formula $A = I^2 R \Delta t$ for series-connected conductors since in this case the current is the same in all the conductors. For parallel connection, the formula $A = U^2 \Delta t / R$ is more convenient since the voltage across the conductors is the same in this case.

The law determining the amount of heat liberated by a current-carrying conductor to the surroundings was established experimentally for the first time by the English scientist J. Joule and independently by the Russian scientist F. Lenz. The Joule-Lenz law is formulated as follows: the amount of heat liberated by a current-carrying conductor is equal to the product of the squared current, the resistance of the conductor, and the time during which the current flows through the conductor:

$$Q = I^2 R \Delta t. \tag{9.17}$$

We have arrived at this law by following the line of reasoning based on the law of energy conservation. Formula (9.17) allows us to calculate the amount of heat liberated in any subcircuit containing various conductors.

Any electrical device (bulb, electric motor, motor, etc.) is rated to consume a certain amount of energy per unit time. Therefore, along with the work of current, the power of current is also a very important concept. The power of current is equal to the ratio of the work of current done over a time Δt to this time interval.

According to the definition of power,

$$P = \frac{A}{\Delta t} = IU. \tag{9.18}$$

This expression for power can be written in several equivalent forms by using Ohm's law for a conductor:

$$P = IU = I^2 R = U^2 / R.$$

The power rating is indicated on most of devices.

?
1. Formulate Ohm's law.
2. What is the temperature dependence of resistance for metals and solutions of electrolytes?
3. What are the main technical difficulties encountered in the practical application of superconductors?
4. Why must the resistance of an ammeter be small and that of a voltmeter high?
5. Give the definition of the work of current.

Problems with Solutions

While solving problems on Ohm's law (9.3), in addition to Ohm's law you must remember that for the series connection the current in all conductors is the same, while the voltage across the subcircuit is equal to the sum of voltages across individual conductors. For the parallel connection, the voltage is the same for all the conductors, while the current in the unbranched circuit is equal to the sum of the currents in the individual conductors. Formulas (9.7), (9.8), (9.10) and (9.12) follow from Ohm's law, but it would be better if you memorize and apply them directly while solving problems.

In solving problems involving the work and power of current, formulas (9.16) and (9.18) should be used.

1. The current in a circuit containing a rheostat is $I = 3.2$ A. The voltage across the rheostat terminals is $U = 14.4$ V. What is the resistance R of the part of the rheostat in which the current flows?

Solution. According to Ohm's law, $I = U/R$, whence $R = U/I$, $R = 4.5\ \Omega$.

2. The device intended for supplying an arbitrary fraction U of a given constant voltage U_0 is known as a voltage divider. It is usually made of a conductor with a high resistivity and is supplied with a sliding contact for a smooth control of voltage U.

Such a voltage divider is called a POTENTIOMETER. The scheme of its connection is shown in Fig. 157. The resistance of the conductor AB is $R_0 = 4$ kΩ, the voltage $U_0 = 220$ V. A load (consumer) having a resistance $R = 10$ kΩ is connected between the sliding contact D and terminal A of the potentiometer.

Find the voltage U across the load when the sliding contact divides conductor AB into two equal parts.

Solution. The circuit contains two series-connected subcircuits: AD, consisting of half the conductor AB (having a resistance $R_0/2$) and the load (of resistance R) connected in parallel; and DB whose resistance is $R_0/2$. The resistance of AD is given by

$$R_{AD} = \frac{\dfrac{R_0}{2}R}{\dfrac{R_0}{2} + R} = \frac{R_0 R}{R_0 + 2R}.$$

The total resistance of the circuit is

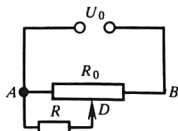$$R_{AB} = R_{AD} + \frac{R_0}{2} = \frac{2R_{AD} + R_0}{2}.$$

Fig. 157

The current in the circuit is

$$I = \frac{U_0}{R_{AB}} = \frac{2U_0}{2R_{AD} + R_0}.$$

The required voltage is given by

$$U = IR_{AD} = \frac{2U_0 R_{AD}}{2R_{AD} + R_0} = \frac{2U_0 R}{R_0 + 4R} = 100 \text{ B}.$$

Exercise 11

1. Electrons flying to the screen of a cathode-ray tube form an electron beam. What is the direction of current in the beam?
2. Why was a wooden stick used instead of a metallic rod in experiments on the potential drop along a conductor described in Sec. 9.2?
3. Determine the cross-sectional area and the length of a copper conductor if its resistance is 0.2 Ω and mass is 0.2 kg. The density of copper is 8900 kg/m³.
4. A galvanometer whose internal resistance is 600 Ω is shunted by a resistance of 25 Ω. What is the change in the scale division of the instrument?
5. A voltmeter whose internal resistance is $R_v = 1000$ Ω is rated for a maximum voltage $U_0 = 150$ V. A sixfold higher voltage of 900 V has to be measured with the help of this voltmeter. What series resistance must be connected to the instrument for this purpose?
6. A voltage of 36 V is applied across the ends of a copper conductor whose length is 300 m. Find the average velocity of ordered motion of electrons in the conductor, if the number density of conduction electrons in copper is 8.5×10^{28} m⁻³.
7. How can we make sure that constant current is actually established in a circular superconductor?
8. A long wire with a constant voltage maintained across its ends has become red hot. A part of the wire is immersed in cold water. Why is the part of the wire remaining above water heated to a higher temperature?
9. The spiral of an electric hot plate has been burnt and becomes shorter after the burnt ends have been connected. What is the change in the amount of heat liberated by the hot plate during a certain time?
10. A hot plate connected to a d.c. circuit has liberated heat Q over a certain time interval. What amount of heat will be liberated by two such hot plates connected to the same circuit in series? in parallel?
11. The aluminium winding of an electromagnet consumes 5-kW power at 0 °C. What amount of power will be consumed if the temperature of the winding has increased to 60 °C during the operation at the same voltage? Solve the problem for the case when the same current is maintained in the winding.
12. To print three-million copies of this textbook, 1200 tonnes of paper is required. This involves an expenditure of 840 000 kWh of electric energy. The scale of these figures can be estimated with the help of the following example.

An electric locomotive pulling the train of such a mass and consuming the same amount of energy during 7 days will cover a distance of 7500 km.

What is the efficiency of the locomotive and the current in its engine, if the voltage in the contact circuit is 3000 V, and the tractive force of the locomotive constitutes about 1/30 of the weight of the train?

9.9. Electromotive Force

If two metal balls bearing unlike charges are connected by a conductor, the electric field of these charges will give rise to a current in the conductor (Fig. 158). This current, however, exists for a very short time. The charges rapidly neutralize each other, the potential difference between the balls vanishes, and the electric field strength becomes zero.

EXTRANEOUS FORCES. To obtain a constant current, it is necessary to maintain a constant voltage between the balls. For this purpose, a device (current source) is needed to carry the charges from one ball to another in the direction opposite to the direction of forces exerted by the electric field of the balls on the charges. In such a device, the forces of nonelectrostatic origin must act in addition to electric forces (Fig. 159). The electric field of charged particles alone (Coulomb's field) cannot maintain a constant current in the circuit.

Any forces acting on electrically charged particles, except for the forces of electrostatic origin (Coulomb's forces), are known as extraneous forces.

The conclusion about extraneous forces required for maintaining a constant current in a circuit becomes even more evident if we turn to the law of energy conservation. The electrostatic field is a potential field. When charged particles move along a closed path, the work of this field is zero. It is well known that the passage of current in conductors is accompanied by the liberation of energy (conductors are heated). Consequently, any circuit must contain a source which supplies energy to the circuit. In addition to
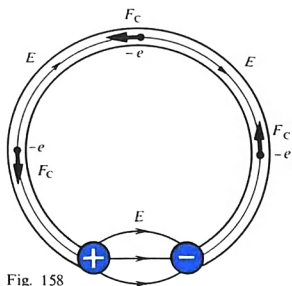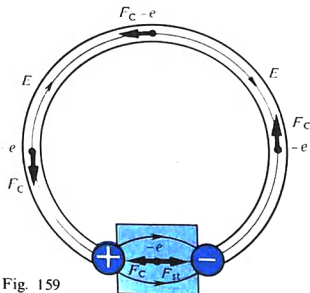

Fig. 158

Fig. 159

179

Coulomb's forces, extraneous (nonpotential) forces must operate in this source. The work of these forces along a closed path must differ from zero. As a result of work done by these forces on charged particles, the latter acquire an energy inside the current source and then give it away while moving in the conductors of the electric circuit.

Extraneous forces are responsible for the motion of particles inside all current sources, viz. generators in electric power plants, voltaic cells, accumulators, etc.

When the circuit is closed, an electric field is generated in all the conductors of the circuit. Inside the source, the charges move under the action of extraneous forces against Coulomb's forces (negative charges move from the plus to the minus pole), while in the remaining circuit they are moved by the electric field (Fig. 159).

THE ANALOGY BETWEEN ELECTRIC CURRENT AND FLUID FLOW. For a better understanding, let us consider the analogy between the electric current in a conductor and the fluid flow in tubes. On any segment of a horizontal tube, the fluid flow is due to a pressure drop across a segment of the tube. The fluid moves towards the decreasing pressure region. But the force of pressure in liquid is a type of elastic forces which are potential forces like Coulomb forces. Therefore, the work done by these forces along a closed path is zero, and these forces alone cannot cause a prolonged circulation of fluid in the tubes. A fluid flow is accompanied by energy losses due to friction. To make a fluid circulate, a pump is needed. The piston of this pump acts on the fluid particles and creates a constant pressure difference between the inlet and the outlet of the pump (Fig. 160). As a result, the fluid flows in the tube. The pump is an analogue of the current source, and the role of extraneous forces is played by the force exerted on the fluid by the moving piston. Inside the pump, the fluid flows from the regions at a lower pressure to the regions at a higher pressure.

THE ORIGIN OF EXTRANEOUS FORCES. Extraneous forces may be of various types. The extraneous force in the generators of an electric power plant is the force exerted by the magnetic field on electrons in the moving conductor (this force was briefly described in the course of "Junior Physics").
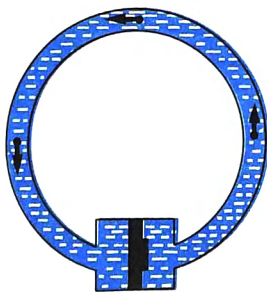

Fig. 160

In a galvanic cell, e.g., a voltaic cell, chemical forces of electromagnetic nature operate. A voltaic cell consists of zinc and copper electrodes placed in a sulphuric acid solution. Chemical forces cause the dissolution of zinc in the acid. Positive zinc ions go over to the solution, and the zinc electrode acquires a negative charge. Copper is dissolved in sulphuric acid to a very small extent. A potential difference appears between the zinc and copper electrodes, which ensures current in the closed electric circuit.

ELECTROMOTIVE FORCE. The action of extraneous forces is characterized by an important physical quantity known as ELECTROMOTIVE FORCE (abbreviated as emf). The electromotive force in a closed circuit is equal to the ratio of the work done by extraneous forces to move a charge along the circuit to this charge:[1]

$$\mathscr{E} = \frac{A_{\text{ex}}}{\Delta q}. \tag{9.19}$$

Like the potential difference, electromotive force is expressed in volts.

We can consider electromotive force on any segment of a circuit. This is the work per unit charge done by extraneous forces not in the entire circuit but only on the given segment. The electromotive force of a galvanic cell is the work done by extraneous forces in transporting a unit positive charge within the cell from one terminal to the other. It should be emphasized once again that the work of extraneous forces cannot be expressed in terms of potential difference, since extraneous forces are nonpotential forces, and the work done by them depends on the shape of the path. For example, the work of extraneous forces in transporting a charge between the terminals of a current source outside it is zero.

9.10. Ohm's Law for a Closed Circuit

Let us consider a simple complete (closed) circuit containing a current source (galvanic cell, accumulator or generator) and a resistor of resistance R (Fig. 161). The current source has an emf $\mathscr{E}$ and a resistance r. The resistance of a current source is often called internal resistance in contrast to the external resistance R of the circuit. In a generator, r is the resistance of the windings, while in a galvanic cell this is the resistance of the electrolyte solution and electrodes.

For a closed circuit, Ohm's law connects the current in the circuit, emf and the total resistance $R + r$ of the circuit. This relation can be established theoretically if we make use of the Joule-Lenz law (9.17) and the energy conservation law.

Suppose that a charge Δq passes during a time Δt through the cross section of a conductor. Then work done by extraneous forces in displacing the charge Δq can be written as $A_{\text{ex}} = \mathscr{E} \, \Delta q$. According to the definition of

---

[1] As can be seen from the definition, emf is numerically equal to the work done in moving a unit positive charge, and not to a force in the usual sense of the word. Though not very exact, the term emf is universally used in literature.

the current (9.1), $\Delta q = I \Delta t$. Therefore,

$$A_{ex} = \mathscr{E} I \Delta t. \tag{9.20}$$

When this work is done over the external and internal subcircuits having resistances R and r, a certain amount of heat is liberated. According to the Joule-Lenz law, this heat is

$$Q = I^2 R \Delta t + I^2 r \Delta t. \tag{9.21}$$

In accordance with the energy conservation law, $A = Q$. Equating (9.20) and (9.21), we obtain

$$\mathscr{E} = IR + Ir. \tag{9.22}$$

The product of current and resistance of a segment of the circuit is called the POTENTIAL DROP ON this SEGMENT. Thus, the emf is equal to the sum of the potential drops on the external and internal subcircuits.

Ohm's law for a closed circuit is usually written in the form

$$I = \frac{\mathscr{E}}{R + r}. \tag{9.23}$$

The current in a closed circuit is equal to the ratio of the emf of the circuit to its total resistance.

The strength of current depends on three quantities: the emf $\mathscr{E}$, and resistances R and r of the external and internal subcircuits. The internal resistance of the current source does not have a noticeable effect on the current if it is small in comparison with the resistance of the external subcircuit ($R \gg r$). In this case, the voltage across the terminals of the source is nearly equal to the emf: $U = IR \approx \mathscr{E}$.

In short-circuiting ($R \rightarrow 0$), the current in the circuit is determined just by the internal resistance of the source, and for an electromotive force of several volts, the current can be very large if r is low (for example, the resistance of an accumulator is $r \simeq 0.1$-0.001 Ω). The wires may melt and the current source may be damaged.

If a circuit contains several series-connected cells having emf's $\mathscr{E}_1$, $\mathscr{E}_2$, $\mathscr{E}_3$, etc., *the total emf of the circuit is equal to the algebraic sum of the emf's of individual cells*. To determine the sign of the emf of any cell, the positive
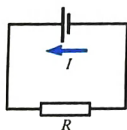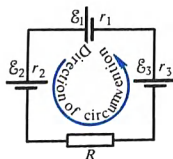

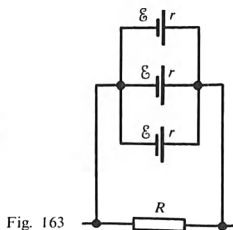Fig. 161   Fig. 162   Fig. 163

direction of circumvention of the circuit should first be chosen conventionally. In Fig. 162, the anticlockwise direction is (arbitrarily) chosen as the positive direction of circumvention of the circuit.

If during the circumvention of the circuit we go over from the negative pole of the source to the positive one, then emf $\mathscr{E} > 0$. Inside the source, extraneous forces do positive work in this case. On the other hand, if we go over from the positive to negative pole during the circumvention of the circuit, the emf is assumed to be negative. Then extraneous forces do a negative work inside the source. For example, for the circuit circumvented in the anticlockwise direction as shown in Fig. 162 we have

$$\mathscr{E} = \mathscr{E}_1 + \mathscr{E}_2 + \mathscr{E}_3 = |\mathscr{E}_1| - |\mathscr{E}_2| + |\mathscr{E}_3|.$$

If $\mathscr{E} > 0$, then according to (9.23), the current $I > 0$, i.e. the direction of current coincides with the direction of circumvention. If $\mathscr{E} < 0$, then the direction of current is opposite to the direction of circumvention of the circuit. The total resistance R_t of the circuit is equal to the sum of all resistances: $R_t = R + r_1 + r_2 + r_3$.

In the case of parallel connection of galvanic cells (or other sources) with equal emfs the emf of the battery equals that of one of the elements (Fig. 163). The internal resistance of the battery is determined using the rule of parallel connection of conductors. The current in the circuit shown in Fig. 163, by Ohm's law for a closed circuit, is calculated by the formula:

$$I = \frac{\mathscr{E}}{R + r/3}.$$

?
1. Why is the electric field of charged particles (Coulomb's field) unable to maintain a constant electric current in the circuit?
2. Which forces are called extraneous?
3. Give the definition of electromotive force.
4. Formulate Ohm's law for a closed circuit.
5. How is the sign of the emf determined in Ohm's law for a closed circuit?

Problems with Solutions

To solve problems involving emf, you must know Ohm's law for a closed circuit (9.23) and be able to determine correct signs for emfs for circuits containing several current sources. Besides, Ohm's law for a conductor (9.3) and expressions (9.16) and (9.18) for the work and power of current should be used in some problems.

1. An accumulator having an emf $\mathscr{E} = 6.0$ V and an internal resistance $r = 0.1$ Ω is connected to an external circuit whose resistance $R = 12.4$ Ω. What amount of heat is liberated in $t = 10$ min in the entire circuit?

Solution. According to Ohm's law for a closed circuit, the current in the circuit is $I = \mathscr{E}/(R + r)$. The amount of heat liberated in the external subcircuit is $Q_1 = I^2Rt$, and in the internal subcircuit, $Q_2 = I^2rt$. The total amount of heat is

$$Q = Q_1 + Q_2 = I^2 (R + r) t = \frac{\mathscr{E}^2 t}{R + r} \simeq 1.7 \text{ kJ}.$$

2. A galvanic cell having an emf $\mathscr{E} = 5.00$ V and an internal resistance $r = 0.2$ Ω is connected to an external resistance $R = 40.0$ Ω. What is the voltage U across this resistance?

Solution. According to Ohm's law for a conductor, $U = IR$. The current in a closed circuit is $I = \mathscr{E}/(R + r)$. Hence $U = \mathscr{E} R/(R + r) \simeq 4.97$ V.

Exercise 12

1. What is the voltage across the terminals of a galvanic cell having an emf $\mathscr{E}$ if the circuit is disconnected?
2. What is the short-circuiting current for an accumulator having an emf $\mathscr{E} = 12$ V and an internal resistance $r = 0.01\,\Omega$?
3. The battery of a pocket lamp is connected to a rheostat. When the resistance of the rheostat is 1.65 Ω, the voltage across it is 3.30 V, while for the resistance of 3.50 Ω, the voltage is 3.50 V. Calculate the emf and the internal resistance of the battery.
4. Galvanic cells having emf's of 4.50 and 1.50 V and internal resistances of 1.50 and 0.50 Ω are connected to a circuit shown in Fig. 164 supplying power to a pocket lamp. What power is consumed by the lamp if the resistance of its filament in the redhot state is 23 Ω?
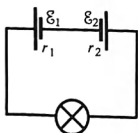

Fig. 164

5. A closed circuit contains a current source having an emf of 6 V and an internal resistance of 0.1 Ω. Plot the graphs of current and voltage in the circuit as functions of the external resistance.
6. Two cells having identical emf's of 4.1 V and identical internal resistances of 4 Ω are connected at like poles. What must be the emf and the internal resistance of a cell that can replace this battery?

Main Points of Chapter 9

A directed (ordered) motion of charged particles is called an electric current. The current is the ratio of the charge Δq transferred through the cross section of a conductor in time Δt to this time interval: $I = \Delta q/\Delta t$.

The current is measured in amperes. According to Ohm's law for a conductor,

$$I = U/R,$$

where U is the voltage across the conductor and R its resistance. The resistance is expressed in ohms: $1\Omega = 1$ V/1 A.

The resistance of metallic conductors increases with temperature almost linearly. At temperatures close to absolute zero, a number of pure metals and alloys completely lose their resistance. This phenomenon is known as superconductivity.

In the ordered movement of charged particles in a conductor, the electric field does a work called the work of current. The work of current over a time Δt on a segment of the conductor is given by

$$A = IU \, \Delta t.$$

The amount of heat liberated in a current-carrying conductor is determined by the Joule-Lenz law:

$$Q = I^2 R \, \Delta t.$$

The power of current is $P = A/\Delta t = IU$.

In any circuit, there must be extraneous forces of nonpotential nature. They act inside a current source connected to the circuit. The ratio of the work done by the extraneous forces to carry a charge Δq along a closed circuit to this charge is called the electromotive force:

$$\mathscr{E} = \frac{A_{\text{ex}}}{\Delta q}.$$

The current in a closed circuit is equal to the ratio of the emf of the circuit to its total resistance:

$$I = \mathscr{E}/(R + r).$$

10 ELECTRIC CURRENT IN VARIOUS MEDIA

10.1. Electric Conductivity of Various Substances

In the two previous chapters, we have considered in detail the electric current in metallic conductors and analyzed the experimentally established voltage-current characteristic for these conductors, i.e. Ohm's law. Metallic conductors are widely used for transmitting energy from current sources to consumers. Besides, these conductors are employed in generators, electric heating devices, and so on.

Along with metals, good conductors (i.e. the *substances containing large numbers of free charged particles*) include aqueous solutions or melts of electrolytes and ionized gases, viz. plasma. These conductors have also found a wide application in engineering. In vacuum electronic devices, electric current is formed by electron flows.

In addition to conductors and dielectrics (viz. *substances containing small amounts of charged particles*), there exists a group of substances whose conductance occupies an intermediate position between conductors and dielectrics. These materials do not conduct current sufficiently well to be called conductors and are not as poor conductors to be called dielectrics. For this reason, they are called SEMICONDUCTORS.

Until recently, semiconductors did not play any significant role in practice. In electrical and radio engineering, only various conductors and dielectrics were used. The situation changed radically (we can even say that a revolution took place in radio engineering) when the possibility to easily control the electric conductivity of semiconductors was first predicted theoretically and then discovered experimentally.

Naturally, conductors are being used as before for energy transmission through wires. Semiconductors are used as elements for current conversion in radio sets, computers, etc.

In this chapter, we shall deal with physical processes accompanying the passage of current through various media. We shall start with metallic conductors. The voltage-current characteristic for these conductors is familiar to us, but nothing has been said so far about its explanation provided by the molecular kinetic theory.

10.2.   Electron Conductivity of Metals

The carriers of free charge in metals are electrons. The electron number density is high (of the order of 10^{28} $1/m^3$). These electrons take part in random thermal motion. Under the action of an electric field, they begin an ordered movement at an average velocity of the order of 10^{-4} m/s (see Sec. 9.1).

EXPERIMENTAL PROOF OF THE EXISTENCE OF FREE ELECTRONS IN METALS. The fact that conductivity of metals is due to the motion of free electrons was proved in the experiments carried out by L. I. Mandelshtam and N. D. Papaleksi (1913) and by Stewart and Tolman (1916).

Schematically, these experiments were carried out as follows. A wire was wound on a coil and its two ends were soldered to two metallic discs insulated from each other (Fig. 165). A galvanometer was connected to the discs with the help of sliding contacts.

The coil is rotated at a high speed, and then abruptly stopped. After the coil has abruptly come to a halt, free charged particles continue to move by inertia for a certain time, and hence an electric current appears in the coil. This current exists for a short time since charged particles are decelerated due to the resistance of the conductor, and the ordered motion of particles forming the electric current ceases.

Observations revealed that a current actually exists in the circuit after the coil has been stopped. Its direction indicates that it is formed by negatively charged particles. The charge transferred in this case is proportional to the ratio of the charge of the particles forming the current to their mass, i.e. to $|q_0|/m$. Therefore, by measuring the charge passing through the galvanometer during the time of the existence of the current in the circuit, the ratio $|q_0|/m$ was determined. Its value was found to be 1.8×10^{11} C/kg. This quantity coincides with the charge-to-mass ratio e/m for the electron, obtained earlier from other experiments.

Fig. 165

MOTION OF ELECTRONS IN A METAL. Under the action of a constant force exerted by the field on the electrons, the latter acquire a certain velocity of ordered motion. This velocity does not increase further with time since a certain retarding force is exerted on the electrons by the ions of the crystal lattice. This force is similar to the resistance acting on a stone sinking in water. As a result, the average velocity of ordered motion of electrons is proportional to the electric field strength in the conductor ($v \propto E$), and hence to the potential difference across the conductor since $E = U/l$, where l is the conductor length.

It is well known that the current in a conductor is proportional to the velocity of ordered motion of particles (see formula (9.2)). Therefore, we can say that the current is proportional to the potential difference across the conductor: $I \propto U$. This is the essence of the qualitative explanation for Ohm's law on the basis of the electron theory of conductivity in metals.

It is impossible to construct a satisfactory quantitative theory of electron motion in a metal with the help of the laws of classical mechanics. As a matter of fact, the conditions of motion of electrons in metals are such that Newton's classical mechanics is inapplicable for describing this motion. This can be clearly seen from the following example. If we determine experimentally the mean kinetic energy of thermal motion of electrons in a metal at room temperature, and calculate the temperature corresponding to this energy with the help of the formula $m\overline{v^2}/2 = (3/2)kT$, the result will be 10^5-10^6 K. This is the temperature of matter in stars. The motion of electrons in a metal obeys the laws of quantum mechanics.

?
1. List the substances which are good conductors of electric current.
2. In the experiment depicted in Fig. 165, the coil was rotated clockwise and then abruptly stopped. Determine the direction of the electric current in the coil at the moment of deceleration.
3. What is the dependence of the velocity of ordered motion of electrons in a metallic conductor on the voltage across the conductor?

10.3. Electric Current in Liquids

Liquids, like solids, may be dielectrics, conductors or semiconductors. Dielectric liquids include distilled water, while among conductors there are electrolyte solutions (acids, alkali and

salts). Liquid semiconductors include melted selenium and melts of sulphides.

ELECTROLYTIC DISSOCIATION. You know from the course of inorganic chemistry why aqueous solutions of electrolytes conduct electric current. When an electrolyte is dissolved, the electric field of polar molecules of water causes the dissociation of electrolyte molecules into ions. This process is known as ELECTROLYTIC DISSOCIATION.

The degree of dissociation, i.e. the fraction of molecules of the dissolved substance which are dissociated into ions, is determined by the temperature, the solution concentration, and the dielectric permittivity ε of the solvent. As the temperature rises, the degree of dissociation becomes higher, and hence the concentration of positively and negatively charged ions increases.

Oppositely charged ions may meet again to form neutral molecules, i.e. undergo a recombination (reunion). Under steady-state conditions, a dynamic equilibrium sets in the solution, during which the number of molecules dissociating into ions per second is equal to the number of pairs of ions which recombine during the same time into neutral molecules.

ION CONDUCTIVITY. The carriers of charge in aqueous solutions and melts of electrolytes are positively and negatively charged ions.

If a vessel containing an electrolyte is connected to an electric circuit, the negative ions start to move towards the positive electrode (anode) and positive ions, to the negative electrode (cathode). As a result, an electric current is established. Since charge is transferred in aqueous solutions and melts of electrolytes by ions, this type of conductivity is called ION conductivity.

Liquids may exhibit electron conductivity as well. This type of conductivity is observed, for example, in liquid metals.

ELECTROLYSIS. In the case of ion conductivity, the passage of electric current is associated with the transport of matter. The substances constituting the electrolyte are deposited on the electrodes. Negatively charged ions give away their extra electrons at the anode (this reaction is known as oxidation in chemistry), while positive ions receive missing electrons at the cathode (reduction reaction). The process of deposition of substances on the electrodes, associated with redox reactions, is known as ELECTROLYSIS.

APPLICATIONS OF ELECTROLYSIS. Electrolysis is widely used in engineering for various purposes. The surface of one metal can be electrolytically coated by a thin layer of another metal (nickel plating, chrome plating, coppering, etc.). This durable coating protects the surface against corrosion.

If special measures are taken for separating the electrolytic coating from the surface on which the metal has been deposited (this can be done, for example, by covering the surface with graphite), a copy of the surface having a complex relief can be obtained.

In printing trade, such copies (stereotypes) are obtained from dies (prints of typesetting on a plastic material). For this, a thick iron layer (or some other material) is deposited on the dies. This makes it possible to reproduce the typesetting with the required number of copies. Earlier, the number of copies of a book was limited to the number of prints that could be obtained

from a typeset (the latter is worn out in printing). Nowadays, the utilization of stereotypes makes it possible to increase the number of copies significantly.

As a matter of fact, stereotypes can be obtained nowadays by the electrolytic method only for books manufactured by high-quality printing.

The process of obtaining removable coatings (galvanoplastics) was designed by the Russian scientist B. S. Yakobi who applied this method for preparing hollow figures for the Isaac cathedral in Leningrad.

Electrolysis is employed for the purification of metals. For instance, unpurified copper obtained from the ore is cast in the form of thick sheets which are then placed in an electrolytic cell as anodes. During electrolysis, the copper anode is dissolved, while impurities including valuable and rare metals precipitate at the bottom of the cell. Pure copper is deposited on the cathode.

With the help of electrolysis, aluminium is obtained from the melt of bauxites. It was this method of production that made aluminium cheap and the most widely used metal (along with iron) in everyday life and engineering.

10.4   Laws of Electrolysis

In electrolysis, a substance is deposited on electrodes. What determines the mass of a substance deposited during a certain time Δt? This mass is obviously equal to the product of the mass m_{0i} of an ion by the number N_i of ions reaching the electrode during the time Δt:

$$m = m_{0i}N_i. \tag{10.1}$$

According to formula (1.5), the mass of an ion is

$$m_{0i} = \frac{M}{N_A}, \tag{10.2}$$

where M is the molar (or atomic) mass of the substance and N_A is Avogadro's number, i.e. the number of ions in a mole.

The number of ions reaching the electrode is

$$N_i = \Delta q/q_{0i}, \tag{10.3}$$

where $\Delta q = I \Delta t$ is the charge passing through the electrolyte over the time Δt, q_{0i} is the charge of the ion, which is determined by the valence n of the atom: $q_{0i} = ne$ (e is the magnitude of the elementary charge).

In the dissociation of molecules consisting of monovalent atoms ($n = 1$), singly-charged ions appear. For example, when KBr molecules dissociate, K^+ and Br^- ions are formed. The dissociation of a copper sulphide molecule results in two doubly-charged ions Cu^{2+} and SO_4^{2-}, since atoms of copper are bivalent in this compound ($n = 2$). Substituting expressions (10.2) and (10.3) into (10.1) and considering that $\Delta q = I \Delta t$ and $q_{0i} = ne$, we obtain
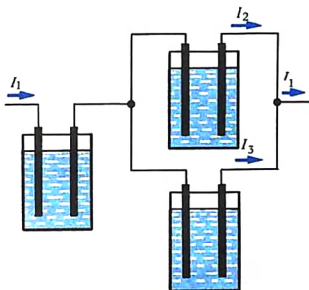
$$m = \frac{M}{neN_A} I \Delta t. \tag{10.4}$$

Fig. 166

FARADAY'S LAW OF ELECTROLYSIS. We denote by k the proportionality factor between the mass m of the substance and the charge $\Delta q = I \Delta t$. Then

$$m = kI \Delta t, \tag{10.5}$$

$$k = \frac{1}{eN_A} \frac{M}{n}, \tag{10.6}$$

where k is the proportionality factor depending on the substance.

Thus, the mass of a substance deposited on the electrodes when the current passes through the electrolyte is proportional to the current and to the time for which it flows.

This theoretically derived statement was first established experimentally by Faraday and is known as FARADAY'S LAW OF ELECTROLYSIS.

Formula (10.5) shows that the coefficient k is numerically equal to the mass of the substance deposited at the electrodes when a charge of 1 C has been carried by ions. The quantity k is called the ELECTROCHEMICAL EQUIVALENT OF A SUBSTANCE and is expressed in kilograms per coulomb (kg/C).

Electrochemical equivalent has a simple physical meaning. Since $M/N_A = m_{0i}$ and $en = q_{0i}$, then according to (10.6) $k = m_{0i}/q_{0i}$, i.e. k is equal to the ratio of the mass of the ion to its charge.

By measuring the quantities m and Δq, we can determine electrochemical equivalents for various substances.

Faraday's law can be verified experimentally. Let us assemble a set-up shown in Fig. 166. All three electrolytic cells are filled with the same electrolyte solution, but the currents passing through them are different. We denote these currents by I_1, I_2 and I_3. Then $I_1 = I_2 + I_3$. By measuring the masses m_1, m_2, and m_3 of substances deposited at the electrodes, we can make sure that they are proportional to the corresponding currents I_1, I_2 and I_3.

DETERMINATION OF THE ELECTRON CHARGE. Formula (10.4) for the mass of a substance deposited at an electrode can be used for

191

determining the electron charge. If follows from this formula that the magnitude of the electron charge is

$$e = \frac{M}{mnN_A} I \Delta t.$$

(10.7)

Knowing the mass m of the substance liberated by the passage of the charge $I \Delta t$, the molar mass M, the valence n of the atoms and Avogadro's number N_A, we can find the magnitude of the electron charge, which turns out to be $e = 1.6 \times 10^{-19}$ C.

The value of the elementary charge was obtained for the first time in 1874 by just this method.

?   1. What phenomenon is known as electrolytic dissociation?
           2. Why does a transport of matter take place during the passage of current through an electrolyte solution while there is no transport of matter when current flows in a metallic conductor?
           3. Formulate Faraday's law of electrolysis.
           4. How can the elementary electric charge be determined from the law of electrolysis?

10.5.   Electric Current in Gases

ELECTRIC DISCHARGE IN A GAS. Let us take an electrometer with two discs of a parallel-plate capacitor connected to it, and impart an electric charge to its rod (Fig. 167). If air is dry enough, the capacitor is not noticeably discharged at room temperature. This indicates that the electric current in air due to the potential difference between the discs is very small. Consequently, the electric conductivity of air at room temperature is very low. Air can be treated as a dielectric.

Let us heat air between the discs by burning a matchstick (Fig. 168). We shall see that the pointer of the electrometer rapidly approaches zero,
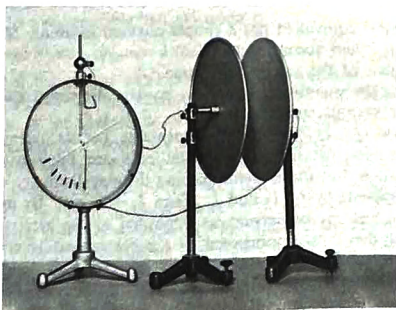
Fig. 167
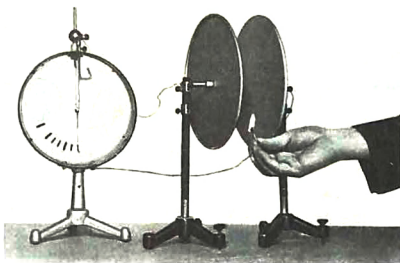
Fig. 168

indicating that the capacitor is discharged. Hence, the heated gas is a conductor, and electric current flows through it.

The passage of current through a gas is called gas discharge.

IONIZATION OF GASES. We saw that air is a very poor conductor at room temperature. During heating, the air conductivity increases. We can also increase the conductivity of air by other methods, e. g., by the action of various types of radiation: ultraviolet, X-ray, radioactive, etc.

Under normal conditions, gases mainly consist of neutral atoms and molecules, and hence are dielectrics. As a result of heating or irradiation, a fraction of atoms is ionized, i.e. the atoms disintegrate into positively charged ions and electrons (Fig. 169). Negative ions can also be formed in a gas. They appear when electrons are combined with neutral atoms.

Ionization of gases upon heating is explained by the increase in molecular velocity with heating. Some molecules acquire such high velocities that upon collisions they partially disintegrate and form ions. The higher the temperature, the larger the number of ions formed.

CONDUCTIVITY OF GASES. The mechanism of conductivity in gases is similar to that for electrolyte solutions and melts. The difference consists in that the negative charge is mainly carried in gases by electrons and not by negative ions as in aqueous solutions and melts of electrolytes.

Thus, in gases, electron conductivity similar to the conductivity in metals is combined with the ion conductivity similar to the conductivity of aqueous solutions and melts of electrolytes. One more difference is of principal importance. In electrolyte solutions, ions are formed as a result of the weakening of intramolecular bonds of ions by the solvent (water) molecules. In gases, ions are generated either as a result of heating or under the effect of external ionizers like radiation.

RECOMBINATION. If an ionizer ceases to act, a charged electrometer will again preserve the charge. This shows that the gas stops being a conductor as soon as the ionizer action is stopped. When all the ions and electrons reach the electrodes, the current ceases.

Moreover, if an electron and a positive ion approach each other, they may again form a neutral atom. This is shown schematically in Fig. 170. Such
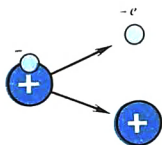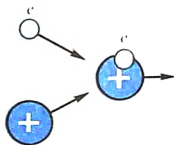
193

Fig. 169

Fig. 170

a process is known as the RECOMBINATION OF CHARGED PARTICLES.

In the absence of an external field, charged particles vanish only as a result of recombination, and the gas becomes a dielectric. If the action of an ionizer is continuous, a dynamic equilibrium is established, in which the number of newly formed pairs of charged particles is on the average equal to the number of pairs disappearing as a result of recombination.

10.6. Nonself-sustained and Self-sustained Discharges

NONSELF-SUSTAINED DISCHARGE. Discharges in a gas at various pressures are studied with the help of a glass tube with two metallic electrodes (Fig. 171).

Suppose that a certain number of pairs of charged particles (positive ions and electrons) is formed in a gas per second with the help of an ionizer.

If the potential difference between the electrodes in the tube is not very high, positive ions move to the negative electrode, while electrons and negative ions move to the positive electrode. As a result, an electric current flows in the tube, i.e. a gas discharge takes place.

Not all the ions formed will, however, reach the electrodes: a fraction of ions recombines to form neutral gas molecules. As the potential difference between the electrodes increases, the fraction of particles reaching the electrodes increases. Hence the current in the circuit increases as well. Finally, a moment comes when all the charged particles formed in the gas per second reach the electrodes during this time, and there is no further increase in the current (Fig. 172). The current is said to attain SATURATION. If the action of the ionizer is terminated, the discharge does not occur any longer since there is no other source of ions. For this reason, such a discharge is called NONSELF-SUSTAINED.

SELF-SUSTAINED DISCHARGE. What will happen to a gas discharge if the potential difference across the electrodes is increased still further?
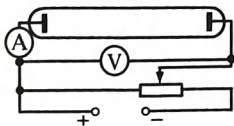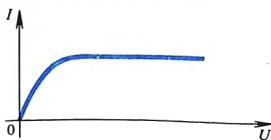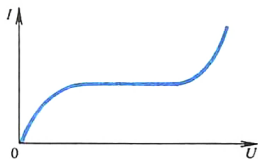

Fig. 171

Fig. 172

194

Fig. 173

Fig. 174

With a further increase in the potential difference, the current would seem to remain unchanged. Experiments show, however, that starting from a certain value of the potential difference between the electrodes, the current increases again (Fig. 173). This means that an additional number of ions besides those formed due to the ionizer appears in the gas. The current may increase by several orders of magnitude, and the number of ions generated in the process of the discharge may become so large that no ionizer is needed any longer for sustaining the discharge. If the external ionizer is removed now, the discharge will not cease. Since no external ionizer is required for sustaining the discharge, the latter is called a SELF-SUSTAINED discharge.

IONIZATION BY ELECTRON IMPACTS. What are the reasons behind a sharp increase in the current at high voltages?

Let us consider a pair of charged particles (a positive ion and an electron) formed due to the effect of an external ionizer. The free electron thus formed starts to move towards the positive electrode (anode) and the positive ion towards the cathode. On its way, the electron encounters ions and neutral atoms. In the interval between two successive collisions the energy of the electron increases at the expense of the work of the external field. Prior to the next collision, the kinetic energy of the electron is proportional to the field strength and the mean free path of the electron (viz. the path between two successive collisions):

$$\frac{mv^2}{2} = eEl. \tag{10.8}$$

If the electron kinetic energy exceeds the work A_i that must be done to ionize a neutral atom, i.e.

$$\frac{mv^2}{2} \geqslant A_i,$$

the collision of the electron with an atom results in ionization. This process is shown schematically in Fig. 174. As a result, two electrons appear instead of one (the impinging electron and the one knocked out of the atom). They, in turn, acquire an energy in the field and ionize the atoms they meet, and so on. Consequently, the number of charged particles rapidly grows, and an electron avalanche is observed. This process is known as IONIZATION BY ELECTRON IMPACTS, or COLLISION IONIZATION. However, the impact ionization alone cannot ensure a self-sustained discharge. Indeed, all the electrons produced in this way move towards the anode, and as soon as they
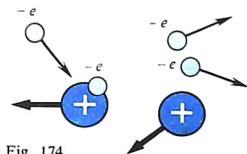
195

reach the anode, the electrons "cease to act". To maintain the discharge, the emission of electrons from the cathode is required. Electron emission may be caused by several factors.

Positive ions formed in collisions of electrons with neutral atoms acquire a high kinetic energy in their motion to the cathode under the action of the field. When such fast ions reach the cathode, they knock electrons out of the cathode surface.

THERMIONIC EMISSION. Besides, a cathode may emit electrons as a result of heating to a high temperature. This process is known as THERMIONIC EMISSION. It can be treated as evaporation of electrons from a metal. In many solid materials thermionic emission occurs at temperatures for which the evaporation of the substance itself is insignificant. Such materials are used for manufacturing cathodes.

In a self-sustained discharge the cathode can be heated due to bombardment by positive ions. If the energy of ions is not very high, the electrons cannot be knocked out of the cathode but appear as a result of thermionic emission.

10.7. Various Types of Self-sustained Discharge and Their Applications

Depending on the properties and state of a gas, the form and arrangement of electrodes, and also on the voltage applied to the electrodes, various types of self-sustained discharge are observed.

GLOW DISCHARGE. Under low pressures (tenths and hudredths of mm Hg), a GLOW DISCHARGE is observed in a tube. To initiate a glow discharge, it is sufficient to apply a voltage of a few hundred volts (and sometimes even lower) to the electrodes. In this case, almost the entire tube (except for a small region at the cathode) is filled by a uniform glow called the positive column (see coloured plate 4).

ELECTRIC ARC. If two carbon rods are brought in contact and a current is passed through them, a large amount of heat is liberated at the contact due to its high resistance. The temperature rises to such an extent that thermionic emission begins. If the carbon electrodes are then moved apart,
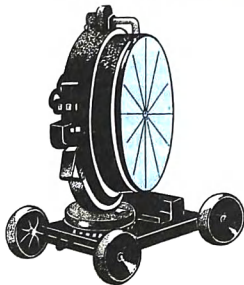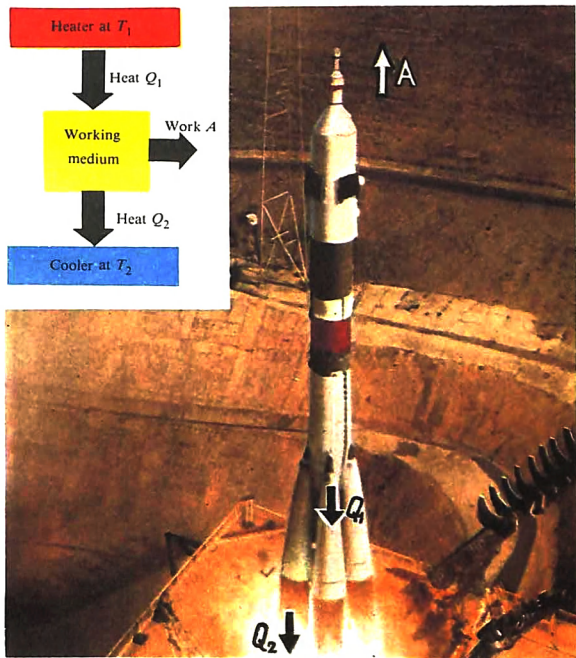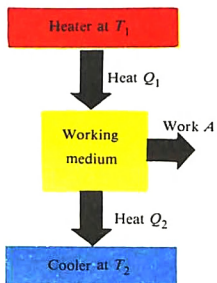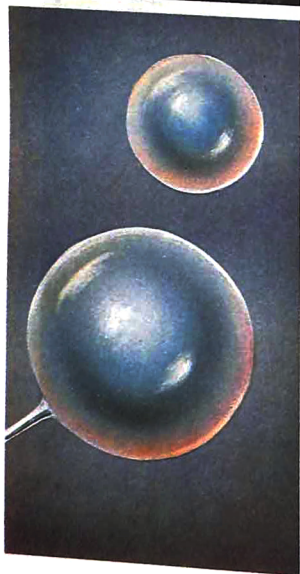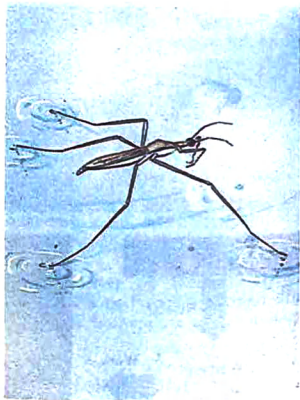

Fig. 175

1. Examples of heat engines: internal combustion engine and rocket engine. During its operation, a heat engine receives an amount of heat Q_1, out of which it gives away Q_2. The balance $A = Q_1 - Q_2$ is spent in doing work.
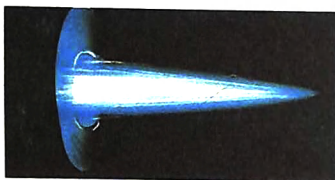
2. Surface tension forces hold a steel needle and a water bug on the water surface. These forces determine the shape and properties of soap bubbles and liquid drops.
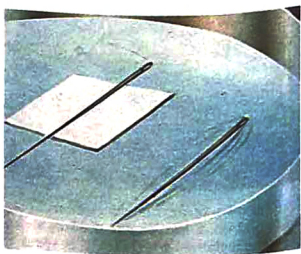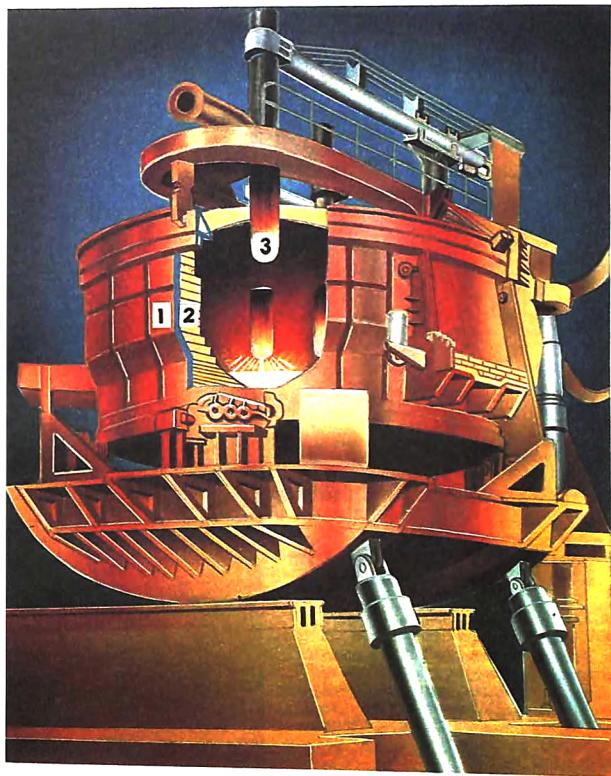
3. An electric arc furnace for melting steel:
1 – jacket, 2 – lining, and 3 – electrode.

4. Examples of gas discharge: a plasma jet emerging from the nozzle of a plasmatron; the glow of the upper strata of the atmosphere, bombarded by charged cosmic-ray particles (aurora borealis); a glow discharge in a tube; an electric arc; a giant spark discharge (lightning), and a corona discharge (St. Elmo's fire).

Fig. 176

Fig. 177

a discharge is struck in the space between them. A column of brightly glowing gas, viz. an ELECTRIC ARC, is observed. In this case, the conductivity of the gas is quite high even under atmospheric pressure since the number of electrons emitted by the negative electrode is very large. In a small arc, the current strength is several amperes, while in large arcs it can be as large as several hundred amperes at a voltage of the order of 50 V.

The electric arc was obtained for the first time in 1802 by the Russian Academician V. V. Petrov.

The arc discharge is a high-power light source used in searchlights (Fig. 175), projectors and cinema projectors.

Electric furnaces used in metallurgy employ the arc discharge as a source of heat (see coloured plate 3). Arc discharge is also used for welding metals (Fig. 176).

OTHER TYPES OF SELF-SUSTAINED DISCHARGE. A discharge whose glowing region resembles a crown (see coloured plate 4) is observed under the atmospheric pressure in the vicinity of sharp parts of a conductor

carrying a large electric charge. This discharge, known as the CORONA DISCHARGE, is caused by a strong electric field (about 3×10^6 V/m) at a charged needle-shaped body. The corona discharge has to be taken into account while dealing with high voltages. When protruding parts of thin wires are present in the circuit, a corona may appear. This leads to a leakage of electric energy. The higher the voltage of a high-tension transmission line, the larger must be the thickness of the wires.

At a high voltage, a SPARK DISCHARGE may appear between the electrodes in air. The spark discharge has the shape of a bundle of bright zigzag bands branching from a thin channel (Fig. 177). This type of discharge appears when the power of a source is insufficient for maintaining an arc or glow discharge. An example of a giant spark discharge is lightning (see coloured plate 4). Lightning appears either between two clouds or between a cloud and the Earth. The current in a lightning reaches 500 000 A, while the potential difference between the cloud and the Earth may be a billion volts.

10.8. Plasma

At very low temperatures, all substances are in the solid state.

Heating causes a transition of a substance to the liquid, and then to the gaseous state.

As the temperature reaches a quite high value, gases are ionized as a result of collisions of rapidly moving atoms or molecules. The substance goes over to a new state of aggregation called plasma[1]. *Plasma is a partially or completely ionized gas in which the densities of positive and negative charges practically coincide.* Thus, a plasma as a whole is an electrically neutral system. Depending on the conditions, the degree of plasma ionization, i.e. the ratio of the number of ionized atoms to their total number, may be different. A completely ionized plasma contains no neutral atoms.

Besides heating, gas ionization and the formation of a plasma can be caused by various types of radiation or bombardment of gas atoms by fast charged particles. In this case, a so-called low-temperature plasma is obtained.

PROPERTIES OF PLASMA. Plasma has some peculiar properties, which allows us to treat it as a special, fourth state of aggregation.

Charged particles of a plasma due to their high mobility easily move under the action of electric or magnetic fields. Therefore, any violation of electric neutrality in separate regions of the plasma caused by the accumulation of like particles disappears quickly. The emerging electric fields move charged particles until electric neutrality is restored and the electric field vanishes.

Unlike a neutral gas where short-range forces operate between the molecules, Coulomb's forces decreasing with increasing distance at a comparatively slow rate, operate among the charged particles of a plasma.

---

[1] From the Greek word *plasma* – moulded. Initially, this word was used in biology for colourless liquid components in blood and live tissues. The word "plasma" acquired a different meaning in physics.

Each particle simultaneously interacts with a large number of neighbouring particles. As a result, in addition to random thermal motion, plasma particles can participate in various ordered (collective) movements. Various types of vibrations and waves can be easily generated in plasmas.

Plasma conductivity increases with the degree of ionization. At a high temperature, a completely ionized plasma attains conductivity close to that of superconductors.

PLASMA IN SPACE. The most part of matter in the Universe (about 99%) is in the state of plasma. The Sun and other stars mainly consist of fully ionized plasma due to their high temperatures.

The interstellar medium filling the space between stars and galaxies also consists of plasma. The density of this medium is very low (less than an atom per cubic centimetre on the average). The atoms of the interstellar medium are ionized by the radiation from stars and by cosmic rays, viz. the beams of fast particles piercing the space in all directions. Unlike the hot plasma of stars, the plasma of interstellar medium has a very low temperature.

Our planet is surrounded by plasma. The upper layer of the atmosphere at a height of 100-300 km, viz. the ionosphere, is an ionized gas. The ionization of the air in the upper layers of the atmosphere is mainly caused by the solar radiation and beams of charged particles emitted by the Sun. Above the ionosphere, there are radiation belts, discovered with the help of artificial satellites. The radiation belts also consist of plasma.

Free electrons in metals possess many properties of plasma. Unlike ordinary plasma, the positive ions in a solid-state plasma cannot move over the entire body.

PRACTICAL APPLICATION OF PLASMAS. Plasma emerges in all types of gas discharge: glow, arc, spark, etc.

The plasma of the positive column of a glow discharge is used in neon lights for advertisements and daylight lamps. In the latter case, the discharge in mercury vapour occurs. The glass tube is coated by a layer of special material – luminophore[1], which glows itself under the action of plasma radiation. A luminophore is chosen in such a way that its glow approaches white light in composition.

The gas-discharge plasma is used in many devices such as gas lasers, viz. quantum generators of light. Lasers are the most powerful sources of light (their operation principle will be described in the next volume of this book).

Plasma jets are employed in magnetohydrodynamic (MHD) generators mentioned on p. 171. The utilization of low-power plasma engines on spacecraft is being considered.

Quite recently, a new device – plasmatron – has been constructed. In this device, powerful jets of high-density plasma are created, which are applied in various fields in engineering: for cutting and welding metals, for drilling holes in hard rocks, etc. (see coloured plate 4). In a plasma jet, many chemical reactions proceed at a higher rate and even those reactions which do not occur under normal conditions may take place.

---

[1] From the Latin word *lumen* meaning "light" and the Greek word *phoros* meaning "carrier".

High-temperature plasma (whose temperature is of the order of 10^7 C) is found to be most promising in the projects associated with the controlled fusion reactions. Intensive investigations are being carried out at present in this direction. As a matter of fact, these reactions are accompanied by the liberation of huge amounts of energy. The realization of this project will give mankind a practically inexhaustible source of energy.

?

1. What is the difference between the dissociation of electrolytes and the ionization of gases?
2. What are the conditions under which a nonself-sustained gas discharge becomes self-sustained?
3. What is thermionic emission?
4. List the main types of self-sustained gas discharge.
5. Give the definition of plasma.

10.9. Electric Current in Vacuum

In cathode-ray tubes and electronic tubes of radio receivers, as well as in many other devices, electrons move in a vacuum. How are the electron beams in a vacuum obtained? What are their properties?

A self-sustained gas discharge in a glass tube with two electrodes (Fig. 171) can occur only under the condition that the gas pressure is not very low. As the pressure is decreased to the value less than 0.0001 mm Hg. the discharge is extinguished, i.e. the current vanishes although the voltage across the electrodes differs from zero.

This is explained by the fact that the amount of atoms is too small to maintain current due to the impact electron ionization and knocking out electrons from the cathode by the ions.

A rarefied gas does not conduct current when its pressure is decreased still further. By continuing the evacuation of the gas from the tube, we may reach such a concentration at which the molecules manage to fly from one end to the other without undergoing a collision with one another. Such a state of the gas in the tube is called a VACUUM.

A rarefied gas becomes a conductor under the action of an ionizer (nonself-sustained gas discharge, see Sec. 10.6). If, however, the amount of gas is so small that we can speak about the state of vacuum, the conductance of space between the electrodes can be ensured only by introducing a source of charged particles in the tube.

The operation of such a source is most often based on the ability of bodies heated to a high temperature to emit electrons. Thermionic emission mentioned in Sec. 10.6 has a very wide range of practical application. In most of modern electronic vacuum devices, the heated cathode is the source of charged particles.

10.10. Two-electrode Electronic Tube (Diode)

As a result of thermionic emission, a heated metallic electrode, unlike a cold electrode, continuously emits electrons which form an "electron cloud" near it. In this case, the electrode acquires a positive charge, and under the action of its electric field, the electrons from the cloud partially return to the electrode. In the equilibrium state, the number of electrons leaving the electrode per second is equal to the number of electrons returning to it during the same time. The higher the temperature of the metal, the higher the density of the electron cloud.

The difference between the hot and cold electrodes soldered into an evacuated vessel leads to unidirectional conductivity of the interelectrode space. When the electrodes are connected to a current source, an electric field emerges between them. If the positive pole of the source is connected to the cold electrode (anode) and the negative to the hot electrode (cathode), the electric field strength is directed towards the hot electrode. Under the action of this field, the electrons partially leave the electron cloud and move towards the cold electrode. The electric circuit is thus closed, and an electric current flows in it. If the polarity of the current source is reversed, the field strength is directed from the cathode to the anode. The electric field repels the electrons of the cloud back to the cathode, and the circuit is disconnected.

DIODE. The property of the interelectrode space to conduct current in one direction is used in electronic devices with two electrodes, viz. VACUUM DIODES.

A modern vacuum diode has the following structure. A cylinder made of glass or cermet and evacuated to an air pressure of 10^{-6}-10^{-7} mm Hg contains two electrodes (Fig. 178a). One of them (cathode) is made in the form of a vertical metallic cylinder normally covered by a layer of oxides of alkali-earth metals (barium, strontium or calcium). Such a cathode is known as oxide-coated. The surface of such a cathode emits upon heating a much larger number of electrons than the surface of a pure metal. Inside the cathode, there is an insulated conductor heated by an alternating current. The hot cathode emits electrons which reach the anode if its potential is higher than that of the cathode.

The anode is a circular or elliptical cylinder coaxial with the cathode. A schematic diagram of a diode is given in Fig. 178b.
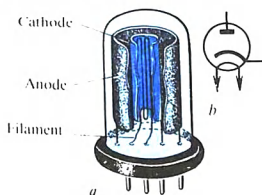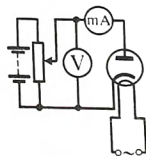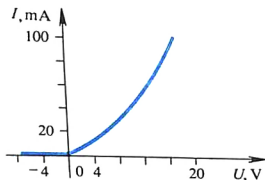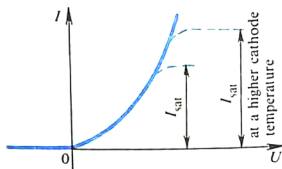

Fig. 178

Fig. 179

Fig. 180

Fig. 181

VOLTAGE-CURRENT CHARACTERISTIC OF A DIODE. The characteristic properties of any electronic device are reflected by its voltage-current characteristic, i.e. the dependence of the current on the potential difference across the terminals of this device. The voltage-current characteristic of a diode can be obtained with the help of the circuit shown in Fig. 179. Unlike the voltage-current characteristic of a metallic conductor, this dependence is not linear (Fig. 180). The main reason behind the nonlinearity of the I-V characteristic of the vacuum diode is that free electrons constituting the current in the interelectrode gap of the diode are emitted by one of the electrodes in limited amounts. Besides, in addition to the field produced by the charges on the electrodes, the field of the spatial charge of the electron cloud near the cathode considerably affects the motion of the electrons.

The higher the voltage between the anode and cathode, the smaller the spatial charge of the electron cloud, the larger the number of electrons reaching the anode, and hence the larger the current in the circuit. If the cathode is not coated by an oxide layer, then at a sufficiently high voltage all the electrons leaving the cathode reach the anode, and a further increase in voltage does not lead to an increase in the current. This means that saturation is attained (dashed line in Fig. 181). If we increase the temperature of the cathode (this can be done by varying the resistance of the rheostat in the filament circuit), a larger number of electrons will leave the cathode. The electron cloud around the cathode will become denser, and the saturation current will set in at a higher voltage between the cathode and anode, its magnitude being larger (the second dashed line in Fig. 181). In a vacuum tube with an oxide-coated cathode saturation cannot be attained since it would require such a high potential difference that the cathode would be destroyed.

Diodes are used for rectifying an alternating electric current.

10.11. Electron Beams. Cathode-ray Tube

If we make a hole in the anode of an electronic tube, a fraction of electrons accelerated by the electric field flies through the hole, forming an electron beam behind the anode. The number of electrons in the beam can be controlled by placing an additional electrode between the cathode and anode and varying its potential. When fast
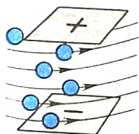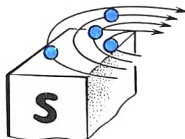
Fig. 182   Fig. 183

electrons of the beam interact with matter, various effects are observed, which are used in practice.

PROPERTIES OF ELECTRON BEAMS AND THEIR APPLICATION. An electron beam incident on bodies causes their heating. In modern engineering this property is used in electron-beam melting of superpure metals in vacuum.

When fast electrons are decelerated in a substance, X-rays are generated. This property is employed in X-ray tubes which will be described in the next volume of this book.

Some materials (like glass, and zinc and cadmium sulphides) glow as a result of bombardment by electrons. At present, this type of materials (luminophores), which convert up to 25% of the electron beam energy into the light energy, are being used.

Electron beams are deflected by the electric field. For example, when electrons pass between the plates of a capacitor, they are deflected from the negative plate to the positive one (Fig. 182).

Electron beams are also deflected in a magnetic field. In their flight over the north pole of a magnet, electrons are deflected to the left, while above the south pole they are deflected to the right (Fig. 183). The deflection of electron flows from the Sun in the magnetic field of the Earth is manifested in the glow of the upper layers of the atmosphere (aurora polaris) that can be observed only at the poles.
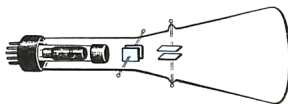

Fig. 185

Fig. 184

The possibility to control an electron beam with the help of an electric or magnetic field and the glow of a luminophore-coated screen under the action of the beam form the basis of the operation of cathode-ray tubes.

CATHODE-RAY TUBE. This tube is the main part of a TV set and an oscillograph, a device for investigating rapidly varying processes in electric circuits (Fig. 184).

A schematic diagram of a cathode-ray tube is shown in Fig. 185. The tube is an evacuated vessel whose one wall serves as a screen. The source of fast electrons, viz. the electron gun, is placed in the narrow part of the tube (Fig. 186). It consists of a cathode, a controlling electrode and an anode (actually, several anodes are arranged one behind another). Electrons emitted by the heated oxide layer from the end face of the cylindrical cathode C surrounded by a heat-insulating screen S. Then they fly through the opening in the cylindrical controlling electrode B (it controls the number of electrons in the beam). Each anode A_1 and A_2 consists of discs with small perforations, inserted into metallic cylinders. A potential difference of hundreds and even thousands volts is created between the first anode and the cathode. The strong electric field accelerates electrons, and they acquire a high velocity. The shape, arrangement and potentials of the anodes are chosen in such a way that the acceleration of electrons is accompanied by beam focussing, i.e. the reduction in the cross-sectional area of the beam almost to a point on the screen.

On its way to the screen, the beam passes successively through two pairs of controlling plates similar to those of a parallel-plate capacitor. If there is no electric field between the plates, the beam is not deflected, and the luminescent point is at the centre of the screen. If a potential difference is applied to the vertical plates, the beam is deflected in the horizontal direction, while the potential difference applied to the horizontal plates deflects the beam in the vertical direction.

The simultaneous operation of the two pairs of plates makes it possible to displace the luminous spot in any direction. Since the mass of the electrons is very small, they respond to a change in the potential difference across the controlling plates almost instantaneously.

In a cathode-ray tube used in a TV set (the so-called picture tube, or kinescope), the electron beam produced by the electron gun is controlled by a magnetic field. This field is produced by the coils mounted on the neck of the tube (Fig. 187).

Cathode-ray tubes have found a very wide application in special devices combined with computers, viz. displays. The information recorded and processed by a computer is displayed on a screen similar to the screen in

Trajectories of electrons
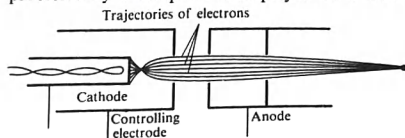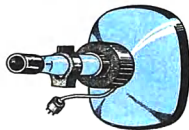

Cathode

Controlling electrode

Anode

Fig. 186

Fig. 187

a TV set. Texts in various languages, the graphs of various processes, the images of real objects, as well as imaginary objects obeying the laws recorded in a computer program can be seen directly on the display.

10.12. Electric Current in Semiconductors

The temperature dependence of electric conductivity of semiconductors makes them completely different in this respect from conductors. Measurements show that for a number of elements (silicon, germanium, selenium, etc.) and compounds (PbS, CdS, etc.), the resistivity does not increase with temperature as for metals (see Fig. 149), but on the contrary, drops very sharply (Fig. 188). Substances exhibiting this property are known as SEMICONDUCTORS.

The curve in Fig. 188 shows that at temperatures close to absolute zero, the resistivity of semiconductors is very high. This means that at low temperatures a semiconductor behaves as a dielectric. As the temperature increases, the resistivity rapidly drops. What is the reason behind this effect?

STRUCTURE OF SEMICONDUCTORS. In order to understand the mechanism of emergence of conductivity in semiconductors, you must know the structure of semiconductor crystals and the origin of the bonds keeping the atoms of a crystal near one another. Let us consider, by way of an example, a silicon crystal.

Silicon is a tetravalent element. This means that there are four electrons on the outer atomic shell, which are rather weakly bound to the nucleus. The number of nearest neighbours of each silicon atom is also four. A two-dimensional diagram illustrating the structure of a silicon crystal is shown in Fig. 189.

A pair of neighbouring atoms interact via an electron-pair bond known as the COVALENT BOND (see textbooks on inorganic chemistry). This bond is formed with the participation of one valence electron from each atom, which are detached from the atoms (are collectivized by the crystal) and spend most
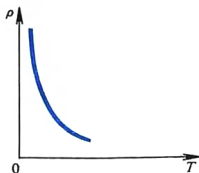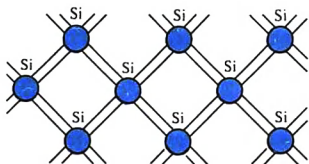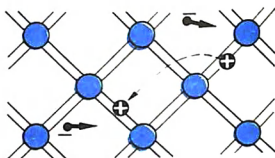

Fig. 188

209

14−441

Fig. 189

Fig. 190

of time in their motion in the space between neighbouring atoms. The positive silicon ions are kept together by the negative charge of these electrons.

It would be wrong to say that a collectivized pair of electrons belongs to only two atoms. Every atom forms four bonds with neighbouring atoms, and any valence electron can move along one of the bonds. When it reaches a neighbouring atom, it may go over to the next atom, and so on. Moving over the entire crystal, valence electrons belong to the crystal as a whole.

Electron-pair bonds in silicon are strong enough and do not rupture at low temperatures. For this reason silicon does not conduct electric current at low temperatures. Valence electrons taking part in the covalent bonds between the atoms are tightly bound to the crystal lattice, and an external electric field does not affect their motion noticeably. A germanium crystal has a similar structure.

ELECTRON CONDUCTIVITY. When a silicon crystal is heated, the kinetic energy of its valence electrons increases, and some of the bonds rupture. Some electrons leave their "habitual route" and become free like electrons in a metal. In an electric field, they travel in the space between crystal lattice sites, forming an electric current (Fig. 190).

The conductivity of semiconductors due to the presence of free electrons in them is called the ELECTRON CONDUCTIVITY. As the temperature rises, the number of ruptured bonds, and hence the number of free electrons, increases. In heating from 300 to 700 K, the number of free charge carriers increases from 10^{17} m^{-3} to 10^{24} m^{-3}.

This leads to a decrease in the resistance.

HOLE CONDUCTIVITY. When a bond is ruptured, a vacancy with a missing electron is formed. It is known as a HOLE. The hole is characterized by an excess positive charge in comparison with the other, normal bonds (Fig. 190).

The position of a hole in a crystal does not remain unchanged. The following process permanently occurs. One of the electrons forming a bond between atoms jumps to fill a hole formed in its neighbourhood and restores the electron-pair bond, a new hole being formed in place of this electron. Thus, a hole may travel over the entire crystal.

If the electric field strength in the sample is zero, the movement of holes equivalent to the motion of positive charges is chaotic and does not produce any current. If, however, an electric field is present, an ordered motion of holes is observed, and thus an electric current associated with the

displacement of holes is added to the electric current due to free electrons. The holes move opposite to the motion of electrons.

Thus, two types of carriers are present in semiconductors, viz. electrons and holes. Therefore, semiconductors exhibit the electron as well as hole conductivity.

We have analyzed the mechanism of conductivity in ideal semiconductors. Such a conductivity is known as the INTRINSIC conductivity of semiconductors.

10.13. Electric Conductivity of Semiconductors Containing Impurities

The intrinsic conductivity of semiconductors is normally not very high, since the number of free electrons is small (for example, in germanium $n_e = 3 \times 10^{13}$ cm^{-3} at room temperature). On the other hand, the number of germanium atoms in 1 cm^3 is 10^{23}. Thus, the number of free electrons constitutes 10^{-10} of the total number of atoms. The intrinsic conductivity of semiconductors is similar in many respects to the conductivity of aqueous solutions and melts of electrolytes. In both cases, the number of free charge carriers increases with the intensity of thermal motion. Consequently, in semiconductors, as well as in solutions and melts of electrolytes, conductivity increases with temperature.

A distinctive feature of semiconductors is that in the presence of impurities, an additional (IMPURITY or EXTRINSIC) CONDUCTIVITY is observed in semiconductors along with the intrinsic conductivity. By varying the impurity concentration, we can noticeably change the number of charge carriers of both signs. Owing to this, it is possible to create semiconductors with a predominant concentration of either positive or negative charge carriers. This property of semiconductors makes them invaluable in many practical applications.

DONOR IMPURITIES. It turns out that the presence of arsenic atoms (even at a very low concentration) increases the number of free electrons many times. This is due to the following circumstance. Arsenic atoms have five valence electrons. Four of them take part in the formation of a covalent bond between a given atom and the neighbouring atoms, say, of silicon. The fifth valence electron turns out to be weakly bound to the atom. It easily leaves the arsenic atom and becomes free (Fig. 191).

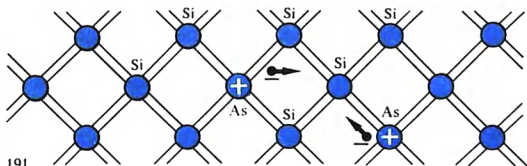When a $1/10^7$ fraction of arsenic atoms is added, the concentration of free

Fig. 191

electrons becomes 10^{16} cm^{-3}. It is 10^3 higher than the free electron concentration in a pure semiconductor.

Impurities which readily give away electrons, and hence which increase the number of free electrons, are called DONOR IMPURITIES.

Since semiconductors containing a donor impurity have a larger number of electrons (in comparison with the number of holes), they are known as n-type semiconductors (from the word "negative"). In n-type semiconductors, electrons are MAJORITY carriers while holes are MINORITY carriers.

ACCEPTOR IMPURITIES. If we use as an impurity indium whose atoms are trivalent, the type of the conductivity in the semiconductor changes. Now an indium atom is missing an electron to form normal electron-pair bonds with the neighbouring atoms. As a result, a hole is formed. The number of holes in the crystal is equal to the number of impurity atoms. Such impurities are known as ACCEPTOR IMPURITIES.

In the presence of an electric field, holes travel along the field, and the hole conductivity emerges. Semiconductors in which the hole conductivity dominates over the electron conductivity are called p-type semiconductors (from the word "positive"). The majority carriers in a p-type semiconductor are holes, while electrons are the minority carriers.

10.14. Electric Current Through a Junction of p- and n-type Semiconductors

Figure 192 represents a schematic diagram of a semiconductor whose right-hand part contains only donor impurities and hence is an n-type semiconductor, while the left-hand part containing acceptor impurities is a p-type semiconductor. Electrons are shown by coloured dots while holes correspond to black spots. The contact between the two semiconductors is known as a p-n JUNCTION.

When a junction is formed, a part of electrons goes over from the n-type to the p-type semiconductor, while holes travel in the opposite direction. The process of diffusion occurring in this case is similar to that observed at the contact of two different gases whose molecules are mixed due to thermal motion. As a result, the n-type semiconductor acquires a positive charge and the p-type semiconductor is charged negatively. Diffusion ceases when the electric field generated in the region of the junction becomes strong enough to prevent a further displacement of electrons and holes.

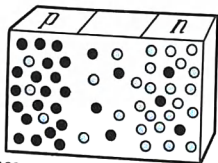Let us connect a semiconductor with a p-n junction to an electric circuit (Fig. 193). We shall first connect the battery in such a way that the p-type
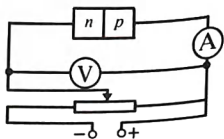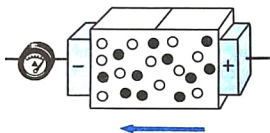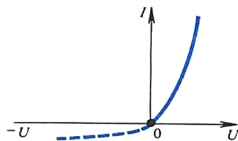

Fig. 192

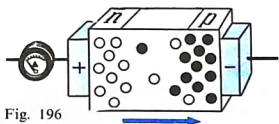Fig. 193

Fig. 194

Fig. 195

Fig. 196

semiconductor has a positive potential and the n-type one has a negative potential. In this case, the current through the p-n junction will be due to the majority carriers: electrons from the n- to the p-zone and holes from the p- to the n-zone (Fig. 194). As a result, the conductivity of the entire sample will be high and its resistance low.

The process considered above is known as the FORWARD CURRENT. The dependence of current on the potential difference, i.e. the voltage-current characteristic for the direct process, is shown in Fig. 195 by the solid line.

Let us change the polarity of the battery. Then at the same potential difference the current will be much weaker than in the direct process. This can be explained as follows. Now the electrons move through the junction from the p- to the n-region, while holes move in the opposite direction. But the number of electrons in the p-type semiconductors as well as the number of holes in the n-type semiconductors is very small. The current through the junction in this case is due to the minority carriers whose number is small (Fig. 196). For this reason, the conductivity of the sample is found to be insignificant and its resistance is high. The so-called BARRIER LAYER is formed. This process is known as the BACKWARD CURRENT. The current voltage characteristic for the backward process is shown in Fig. 195 by the dashed line.

Thus, a *p-n junction is asymmetric with respect to the current: the resistance of the junction in the forward direction is much lower than in the backward direction.* This property of the p-n junction is employed for the rectification of alternating current. During the half-period when the potential of the p-type semiconductor is positive the current flows freely through the junction. During the next half-period, the current through the p-n junction is virtually equal to zero.

213

10.15. Semiconductor Diode

Semiconductor diodes are being used widely in radiocircuits along with two-electrode electronic tubes since they have some significant advantages over electronic devices. Charge carriers in an electronic tube (electrons) are generated as a result of thermionic emission. For this a special current source is required for heating the cathode filament.

In a *p-n* junction, charge carriers appear when an acceptor or donor impurity is introduced into a crystal. Thus, there is no need to use a power source for obtaining free charge carriers. In complex circuits, a considerable amount of energy is saved due to this circumstance. Semiconductor rectifiers have a much smaller size in comparison with vacuum tubes rated for the same current. Therefore, the devices based on semiconductors are more compact.

The advantages of semiconductor devices are of special importance when they are used in the Earth's satellites, spaceships, computers.

Semiconductor diodes are prepared from germanium, silicon, selenium and other materials.

Let us consider a *p-n* junction created in a germanium diode, having a high *n*-type conductivity, by introducing a small amount of a donor impurity. This junction cannot be prepared by mechanically bringing in contact two semiconductors of different types of conductivity since the gap between the semiconductors is too large in this case. On the other hand, the thickness of the *p-n* junction should not exceed interatomic distances. For this reason indium is melted into one of the surfaces of the sample. As a result of the diffusion of indium atoms into the germanium single crystal, a *p*-type region is formed near the surface of germanium. The remaining part of the germanium sample, where indium atoms have not reached, has *n*-type conductivity as before. A *p-n* junction appears between the regions with different types of conductivity (Fig. 197). In a semiconductor diode, germanium acts as the cathode and indium as the anode.

To shield the germanium crystal from the harmful effects of air and light, it is enclosed in a hermetic metallic shell (Fig. 198). Semiconductor rectifiers have a high reliability and a long service life. However, they can operate only in a limited temperature range (from -70 to $125\,°C$).
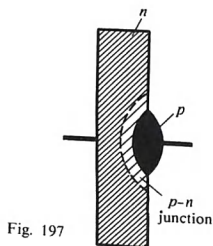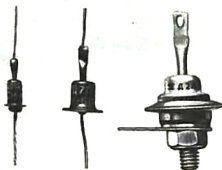

Fig. 197

Fig. 198

214

10.16. Transistors

The properties of the *p-n* junction in semiconductors can be used for the amplification and generation of electric oscillations. In modern radioengineering, SEMICONDUCTOR TRIODES, or TRANSISTORS having various combinations of *p-n* junctions are mainly used.

Let us consider a germanium or silicon transistor with donor or acceptor impurities introduced in a certain way. The distribution of impurities is such that a very thin (of the order of several micrometers) layer of an *n*-type semiconductor is formed between two layers of a *p*-type semiconductor (Fig. 199). This thin layer is called the BASE.

Thus, two *p-n* junctions are formed in the crystal, so that their forward directions are opposite. Three leads from the regions with different types of conductivity make it possible to connect the transistor into a circuit shown in Fig. 199. With such a connection, the left *p-n* junction is DIRECT and separates the base from the *p*-type region called the EMITTER. If the right *p-n* junction were absent, the current in the "emitter-base" circuit would depend on the voltage of the sources (battery *B1* and the source of alternating voltage) and on the resistance of the circuit, including the small resistance of the forward emitter-base junction.

Battery *B2* is connected so that the right *p-n* junction in the circuit shown in Fig. 199 is BACKWARD. It separates the base from the right-hand region with the *p*-type conductivity called the COLLECTOR. If the left-hand *p-n* junction were absent, the current in the collector circuit would be close to zero since the resistance to the backward current is very high. If, however, there exists a current through the left-hand *p-n* junction, a current also appears in the collector circuit, the current in the collector being only slightly smaller than the current in the emitter.

This can be explained as follows. When a voltage is applied between the emitter and the base, the majority carriers in the *p*-type semiconductor, viz. holes, penetrate the base where they are minority carriers. Since the thickness of the base is very small and the number of majority carriers (electrons) in it is not large, the holes appearing in the base almost do not recombine with
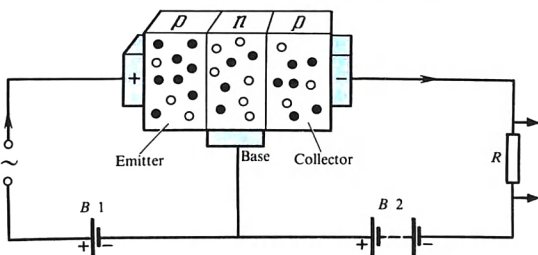

Fig. 199

Fig. 200

the electrons of the base and penetrate the collector as a result of diffusion. The right-hand *p-n* junction is impenetrable for the majority carriers of the base (electrons) but not for holes. The holes are entrained by the electric field in the collector, and the circuit is thus closed. The current branching to the emitter circuit from the base is very small since the cross-sectional area of the base in the horizontal plane (according to Fig. 199) is much smaller than the cross-sectional area in the vertical plane.

The current in the collector, which is virtually equal to the current in the emitter, varies with the current in the emitter. The resistance *R* of the resistor weakly affects the current in the collector and thus can be made sufficiently high. Controlling the emitter current with the help of the source of alternating voltage connected to the circuit, we obtain a synchronous variation of the voltage across the resistor *R*. At a large resistance of the resistor, the change of the voltage across it may exceed the change in the voltage of the signal in the emitter circuit by several orders of magnitude. This corresponds to the voltage amplification. The power liberated at the load *R* will considerably exceed the power consumed in the emitter circuit. Thus, power amplification takes place.

APPLICATION OF TRANSISTORS. Transistors (Fig. 200) have found a very wide application in modern engineering. They replace electronic tubes in many electric circuits in scientific, industrial and domestic appliances. Portable radio receivers in which the semiconductors are used are sometimes called transistors in everyday life. The main advantage of transistors, as well as semiconductor diodes, over electronic tubes is the absence of a hot cathode consuming a considerable amount of power and requiring a certain time for its heating. Semiconductor devices have dimensions and mass of one or even two orders lower than for the corresponding electronic devices. Moreover, they operate at lower voltages.

The drawbacks of transistors are the same as for semiconductor diodes. They are very sensitive to an increase in temperature, strong currents and highly penetrating radiation.

216

10.17. Thermistors and Photoresistors

THERMISTORS. The electric resistance in semiconductors strongly depends on temperature. This property is used for measuring temperature from the current in a circuit containing a semiconductor. Such instruments are known as THERMISTORS or TEMPERATURE-SENSITIVE RESISTORS.

A thermistor is a simple semiconductor instrument. Thermistors are manufactured in the form of rods, tubes, discs, washers or beads, their size ranging from several micrometres to several centimetres (Fig. 201).

The range of temperatures measurable by most of thermistors is from 170 to 570 K. However, there exist thermistors for measuring very high (~1300 K) and very low (4-80 K) temperatures.

Thermistors are used for the remote measurements of temperature, in fire alarms, and so on.

PHOTORESISTORS. The electric conductivity of semiconductors increases not only with heating but also upon irradiation.

This can be verified with the help of a set-up whose schematic diagram is shown in Fig. 202. When the semiconductor is illuminated (Fig. 203), the current in the circuit noticeably increases. This points to an increase in the conductivity (decrease in the resistance) of the semiconductor under the action of light. This effect is not connected with heating since it can be observed at a constant temperature.

The electric conductivity increases as a result of the rupture of the bonds and generation of electrons and holes at the expense of the energy of light incident on the semiconductor. This phenomenon is known as the PHOTOELECTRIC EFFECT.

The devices in which the photoelectric effect in semiconductors is used are called PHOTORESISTORS. The compactness and high sensitivity of photoresistors make them convenient for use in various branches of science and technology where weak light flows must be registered and measured. Photoresistors are used for determining the surface quality, for checking the size of articles, and so on.

?
1. What is the difference in the temperature dependences of resistances of semiconductors and metals?
2. What are the mobile charge carriers in a pure semiconductor?
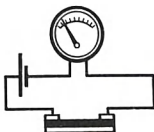3. What happens when an electron meets a hole?

Fig. 201

Fig. 202

Light

Fig. 203

4. Why is the resistance of a semiconductor strongly dependent on the presence of impurities?
5. Which impurity (donor or acceptor) should be introduced for obtaining an *n*-type semiconductor?
6. What are the majority and minority charge carriers in a semiconductor with an acceptor impurity?
7. What are the advantages of a semiconductor diode over a vacuum-tube diode?

Problem with Solution

The quantitative relations for the electric current in metals and electrolytes are very simple. The problems involving Ohm's law which is valid for these conductors were considered in the previous chapter. Here we shall mainly deal with problems where the laws of electrolysis are applied.

A conducting sphere of radius $R = 5$ cm is placed in an electrolytic cell filled with copper sulphate solution. What is the increase in the mass of the sphere, if the deposition of copper lasts for $t = 30$ min, and the electric charge supplied to every square centimetre of the spherical surface per second is 0.01 C? The molar mass of copper M is 0.0635 kg/mole.

Solution. The area of the spherical surface is $S = 4\pi R^2 = 314$ cm^2. Consequently, the charge transported by copper ions during $t = 30$ min $= 1800$ s is $\Delta q = 0.01$ C/(cm$^2\cdot$s) $\times 314$ cm$^2 \times 1800$ s $= 5652$ C. The mass of deposited copper is $m = M\,\Delta q/neN_A = 2 \times 10^{-3}$ kg

Exercise 13

1. A zinc electrode with mass $m = 0.01$ kg is placed in an electrolytic cell intended for zinc-plating of metallic articles. What charge should be passed through the cell to use up the electrode completely? The electrochemical equivalent for zinc is $k = 3.4 \times 10^{-7}$ kg/C.
2. A layer of copper having a mass of 0.316 g has been deposited on the cathode of an electrolytic cell in 10 min at a current of 1.6 A. Determine the electrochemical equivalent for copper.
3. How should the electrodes be arranged for coating electrolytically the inner surface of a hollow metallic body?
4. During nickel-plating of an article lasting for 2 h, the current through the cell was 25 A. The electrochemical equivalent for nickel is 3×10^{-7} kg/C, the density of nickel is 8.9×10^3 kg/m^3. What is the thickness of the nickel layer if the surface area of the article is 0.2 m^2?
5. A uniform electric field of strength $\vec{E}$ is produced in a metal and in vacuum. Is the distance covered by an electron during the same time the same in two cases? The initial velocity of the electron is zero.
6. Determine the velocities of electrons at the exit of an electron gun for the potential differences between the cathode and anode of 500 and 5000 V.

The electric conductivity in metals is due to free electrons (electron conductivity).

The electric conductivity of aqueous solutions of electrolytes is due to positive and negative ions (ion conductivity).

In the case of the ion conductivity, the passage of current is accompanied by the deposition of substances constituting an electrolyte on the electrodes. The process, which is widely used in engineering, is called electrolysis.

The mass of the substance deposited in electrolysis during the time Δt is $m = MI\,\Delta t/enN_A$, where M is the molar mass of the substance, n is its valence, N_A is Avogadro's number, and e is the electron charge.

At temperatures close to room temperature, gases consist of neutral molecules and are dielectrics. Upon heating, as well as under the action of radiation and other factors, ionization of gases takes place and they become conductors.

The conductivity of gases is due mainly to positive ions and electrons. A gas discharge which is terminated as soon as an ionizer is switched off is known as a nonself-sustained discharge.

A discharge is known as self-sustained if it exists without an ionizer. In this type of discharge, ions and electrons are formed as a result of ionization by electron impact (collision), thermionic emission, and so on.

In order to create a current in vacuum, a source of electrons, viz. a hot cathode, should be placed in a vacuum tube.

In modern radio engineering, semiconductors are widely used, whose resistance drops with increasing temperature and strongly depends on the presence of impurities. Since the conductivity of semiconductors can be easily controlled, they are conveniently used in semiconductor diodes and transistors, viz. the devices for the amplification and generation of electric oscillations.

11 MAGNETIC FIELD

11.1. Interaction of Currents. Magnetic Field

The forces acting between stationary electric charges are determined by Coulomb's law. According to the short-range theory this interaction is realized as follows: every charge produces an electric field, the field of one charge acting on another charge, and vice versa.

However, the forces of a different kind may also exist between electric charges. Their action can be observed in the following experiment. Let us take two flexible conductors, fix them in the vertical position and connect their lower ends to the terminals of a current source (Fig. 204). No attraction or repulsion is observed in this case. [1] If, however, we connect the other ends of the conductor by a wire so that opposite currents flow in the conductors (Fig. 205), they start to repel each other. If the direction of current in the conductors is the same the conductors attract each other (Fig. 206).

The interaction between current-carrying conductors, i.e. the interaction between moving electric charges, is called magnetic interaction. The forces with which current carrying conductors act on each other are known as MAGNETIC FORCES. The idea of magnetic interaction was described in the course of "Junior Physics". In the next volume of this book, magnetic interaction will be considered in greater detail.

MAGNETIC FIELD. According to the short-range theory, the current flowing in a conductor cannot act directly on another current.

Like an electric field produced in the space surrounding stationary electric charges, the so-called magnetic field appears in the space surrounding currents.

The electric current in one of the conductors produces in the surrounding space a magnetic field which acts on the current in the other conductor. On the other hand, the field produced by the second current acts on the first current.

The magnetic field is a special form of matter through which the interaction between moving electrically charged particles is realized.

---

[1] The conductors are charged by the current source, but their charges are negligibly small for a few volts potential difference between them. Hence the Coulomb forces are not manifested here.
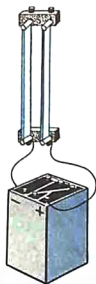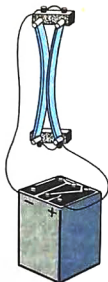
Fig. 204   Fig. 205   Fig. 206

The basic properties of the magnetic field, which have been established experimentally, are listed below.

1. *The magnetic field is generated by an electric current* (moving charges).

2. *The existence of the magnetic field is revealed from its action on a current* (moving charges).

Like the electric field, the *magnetic field really exists independently of us or of our knowledge about it.* An experimental proof of the reality of the magnetic field, as well as of the electric field, is the existence of electromagnetic waves (see Sec. 8.7).

A CLOSED CURRENT LOOP IN A MAGNETIC FIELD. It would appear convenient to take a small element of current, i. e. a small segment of a thin current-carrying conductor, for investigating the magnetic field in the same way as we took a small charged body for investigating the electric field. A continuous current cannot, however, exist in a segment of a conductor since any circuit with direct current must be closed.

The magnetic field can best be studied with the help of a closed loop of a small (in comparison with distances over which the magnetic field varies noticeably) size. For example, we can take a small flat frame of an arbitrary shape, made of a wire (Fig. 207). The leads should be arranged close to each other or braided. Then the resultant magnetic force acting on them will be zero.

The following experiment helps to reveal the nature of action of the magnetic field of a current loop. Let us suspend a small plane loop consisting of several wire turns with the help of braided thin flexible wires. We place a vertical wire at a distance considerably larger than the size of the loop (Fig. 208a). When the current is passed through the wire and the loop, the latter turns and orients itself so that the current-carrying wire is in the same plane with the loop (Fig. 208b). If the current in the wire is reversed, the loop turns through 180°.

It is known from the course of "Junior Physics" that the magnetic field can be created not only by an electric current but also by permanent magnets. If we suspend a current loop between the poles of a magnet by
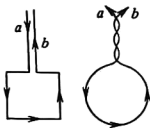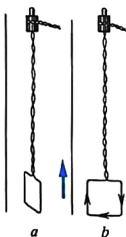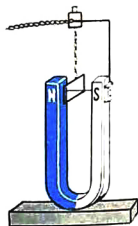
221

Fig. 207

Fig. 208

Fig. 209

flexible wires, it will turn until its plane becomes perpendicular to the line connecting the poles (Fig. 209). Thus, the *magnetic field produces the orienting action on a current loop.*[1]

11.2. Magnetic Induction

The electric field is characterized by a vector, viz. the electric field strength. To characterize the magnetic field, a special physical quantity should be introduced. This quantity is known as the MAGNETIC INDUCTION VECTOR and is denoted by $\vec{B}$.

MAGNETIC NEEDLE. It was shown above that a current loop placed in a magnetic field on a flexible suspension which exerts no elastic force that would prevent the loop orientation turns until it assumes a certain position relative to the field. It is known from the course of "Junior Physics" that a magnetic needle behaves in the same way. The needle is a small elongated magnet with two poles (south S and north N) at the ends.

DIRECTION OF THE MAGNETIC INDUCTION. The orienting action of the magnetic field on a magnetic needle or a current loop can be used to determine the direction of the magnetic induction.

By definition, the direction of the magnetic induction vector is the direction from the south pole S to the north pole N in which the magnetic needle is oriented in a magnetic field. This direction coincides with the direction of the positive normal to the closed current loop (Figs. 210, 211).

The positive normal is directed along the motion of a screw (with the right-hand thread) rotated in the direction of the current in the loop.

Having a current loop or magnetic needle, we can determine the direction of the induction vector at any point of a magnetic field. Figures 212 and 213 illustrate the experiments with a magnetic needle, similar to the experiments with a frame (Figs. 208 and 209).

---

[1] Experiments show that only a uniform magnetic field produces the orienting action on a current loop. In a nonuniform field, the loop is also in translatory motion, being attracted to or repelled from the current-carrying conductor.
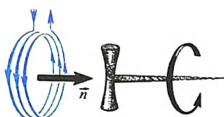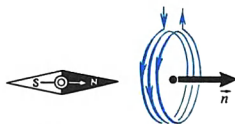
Fig. 210

Fig. 211

The magnetic needle is oriented in the magnetic field of a straight current-carrying conductor along the tangent to the circle (Fig. 213). The plane of the circle is perpendicular to the conductor, its centre lying on the conductor axis. The direction of the magnetic induction vector is established by using the RIGHT-HAND SCREW RULE: IF THE DIRECTION OF THE TRANSLATIONAL MOTION OF THE SCREW COINCIDES WITH THE DIRECTION OF THE CURRENT IN THE CONDUCTOR, THE DIRECTION OF ROTATION OF THE SCREW HANDLE COINCIDES WITH THE DIRECTION OF THE MAGNETIC INDUCTION VECTOR.

The direction of the induction vector of the magnetic field of the Earth is determined each time when a compass is used to find the direction.

11.3. Magnitude of Magnetic Induction

MOMENT OF FORCE ACTING ON A CURRENT LOOP. A uniform magnetic field does not cause the translational motion of a current loop or magnetic needle. According to the laws of mechanics, this means that the vector sum of the forces acting on the loop or needle is zero. But the loop is not in equilibrium unless its normal coincides with the direction of the magnetic induction. This means that the moment of force acting on the loop differs from zero.[1]

---

[1] It should be recalled that the moment of force (or torque) is the product of the force and the arm, viz. the length of the perpendicular dropped from the axis on the line of action of the force. Torques cause rotation of bodies.
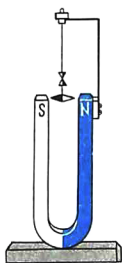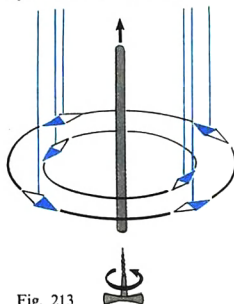

Fig. 212

Fig. 213

The magnitude of this moment does not remain unchanged for a given current in the loop. It depends on the orientation of the plane of the loop relative to the vector $\vec{B}$. When the plane of the loop is perpendicular to the magnetic induction, the moment of force is zero.

The torque acting on a loop can be measured. For this purpose, a device similar to Coulomb's torsion balance is used. A current loop is suspended from a thin elastic wire instead of a rocker with a charged ball. To keep the loop in the initial position after its circuit has been closed, the wire should be twisted through a certain angle. The magnitude of this angle is used to determine the torque acting on the loop. This torque is maximum when the normal to the loop is perpendicular to vector $\vec{B}$.

MAGNITUDE OF MAGNETIC INDUCTION. By varying the current I in the loop and using the loops with different areas S, we can establish an important fact: THE MAXIMUM MOMENT OF FORCE M EXERTED ON THE CURRENT LOOP BY THE MAGNETIC FIELD IS PROPORTIONAL TO THE PRODUCT OF THE CURRENT AND THE AREA OF THE LOOP: $M \propto IS$.

This experimental result can be used for determining the magnitude of the magnetic induction. Indeed, since $M \propto IS$, the ratio M/IS will depend neither on the current in the loop nor on its area. This ratio characterizes the magnetic field at the site where the loop is located.

The magnitude of magnetic induction is the ratio of the maximum torque M acting on the current loop to the product of the current and the area of the loop:

$$B = M/IS. \tag{11.1}$$

The magnetic field is completely characterized by the magnetic induction $\vec{B}$. By measuring the moment of force acting on a current loop, the direction and magnitude of the magnetic induction of the field at each point can be determined.

The relation determining the magnetic induction of a small current element is rather cumbersome and will not be considered here.

UNIT OF MAGNETIC INDUCTION. For the unit of magnetic induction, we take the magnetic induction of a field exerting the maximum torque $M = 1$ N·m on a current loop having an area of 1 m² and a current of 1 A:

1 unit of magnetic induction = 1 (N·m)/(A·m²) = 1 N/(A·m).

The unit of magnetic induction is called the TESLA (T) after the Yugoslavian electrical engineer N. Tesla.

MAGNETOMETER. In the instruments used in practice for measuring magnetic induction, viz. magnetometers, a small magnet is used instead of a current loop. The magnetic induction is judged from the torque exerted on the magnet by the magnetic field.

11.4. Magnetic Field Lines. Magnetic Flux

MAGNETIC FIELD LINES. A visual pattern of a magnetic field can be obtained by plotting the so-called LINES OF MAGNETIC INDUCTION, or magnetic field lines. The magnetic field lines are the lines tangents to which coincide with the vector $\vec{B}$ at a given point of the field (Fig. 214). In this respect, the magnetic field lines are similar to the lines of force of the electric field.

Let us plot the lines of the magnetic field for a straight current-carrying conductor. It follows from the experiments considered above that in the case under consideration, the magnetic field lines are concentric circles lying in a plane perpendicular to the conductor axis (Fig. 215). The arrows on the lines indicate the direction of the magnetic induction vector, viz. the tangent to the line. As in the case of electric field lines, the magnetic field lines are plotted so that their density is proportional to the field at a given region.

Let us also consider the pattern of the magnetic field for a current-carrying coil (solenoid). The pattern of the lines of magnetic induction, plotted with the help of magnetic needles or current loops, is shown in Fig. 216 (the section of the solenoid is drawn). If the length of the solenoid is much larger than its diameter, the field inside the solenoid can be assumed to be uniform. The magnetic lines of such a field are *parallel* and have a *uniform* density.

Using iron filings, the pattern of magnetic field lines can be made visual. This method was described in the course of "Junior Physics".

Every piece of iron put on a carboard sheet is magnetized in a magnetic field and behaves as a small magnetic needle. The availability of a large number of needles allows us to determine the direction of the magnetic field at a large number of points and hence to determine the arrangement of magnetic field lines more accurately. Figures 217-220 represent several patterns of magnetic field lines.
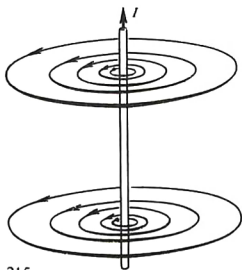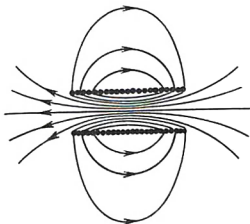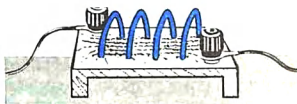
Fig. 214

Fig. 215
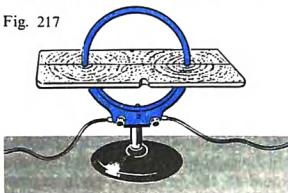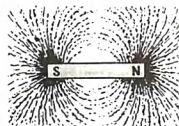
Fig. 216

Fig. 217

Fig. 219

Fig. 218

Fig. 220

VORTEX FIELD. *An important property of magnetic field lines is that they have neither beginning nor end. These lines are always closed.* It should be recalled that the situation with the electric field is quite different. Electric field lines always have a source: they start from the positive charge and terminate at the negative one.

Fields with closed lines of force are termed vortex fields. The magnetic field is a vortex field.

The closed nature of magnetic lines is a fundamental property of the magnetic field. It consists in the fact that the *magnetic field has no source. Magnetic charges similar to electric charges do not exist in nature.*

MAGNETIC FLUX. Magnetic induction $\vec{B}$ characterizes a magnetic field at each point of space. We can introduce another quantity which depends on the magnitude of vector $\vec{B}$ not at a single point but at all points of a surface bounded by a plane current loop.

For this, we consider a closed loop of area S placed in a uniform magnetic field. The normal $\vec{n}$ to the plane of the loop is at an angle α to the direction of the magnetic induction $\vec{B}$ (Fig. 221). The magnetic flux Φ through the surface area S is equal to the product of the magnitude of magnetic induction $\vec{B}$, the area S, and the cosine of the angle α between vectors $\vec{B}$ and $\vec{n}$ (the normal to the surface):

$$\Phi = BS \cos \alpha. \tag{11.2}$$

The product $B \cos \alpha = B_n$ is the projection of the magnetic induction vector onto the normal to the plane of the loop. Therefore,

$$\Phi = B_n S. \tag{11.3}$$

The magnetic flux can be visually interpreted as the quantity proportional to the number of the magnetic field lines piercing a surface of area S.

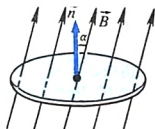The unit of magnetic flux is the WEBER. The magnetic flux of one weber

Fig. 221

(1 Wb) is produced by a uniform magnetic field having an induction of 1 T through a surface which has an area of 1 m² and is perpendicular to the magnetic induction.

11.5. Electrical Measuring Instruments

The orienting effect of the magnetic field on a current loop is used in electrical measuring instruments of magnetoelectric type, viz. ammeters and voltmeters.

A measuring instrument of this type has the following construction. A coil is wound around a light aluminium rectangular frame with a pointer connected to it. The frame is fixed on two half-axles. It is kept in equilibrium position by two thin spiral springs (one of them is shown in Fig. 222). The moment of elastic forces exerted by the springs, which returns the coil to the equilibrium position, is proportional to the angle of deflection of the pointer from the equilibrium position. The coil is placed between the poles of a permanent magnet with pole pieces of special shape.

A cylinder made of soft iron is placed inside the coil. Such a construction ensures the radial direction of the magnetic field lines in the region where the turns of the coil are located (Fig. 223). As a result, in any position of the coil, the moment of force exerted by the magnetic field has the maximum value which remains unchanged for the same current. Vectors $\vec{F}$ represent the forces exerted on the coil by the magnetic field and producing a torque rotating the coil. The current-carrying coil rotates unless the moment of the elastic force exerted by the springs balances the moment of force exerted on the frame by the magnetic field.[1] By doubling the current, we see that the

---

[1] Obviously, the torque acting on a coil with several loops (turns) is proportional to the number N of turns. Indeed, the magnetic field exerts the same action on each loop, and N times larger torque must act on the coil with N turns.
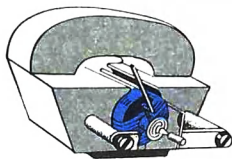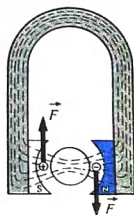

Fig. 222

Fig. 223

pointer turns through a doubly large angle, and so on. This is so because the maximum torque M exerted on the coil by the magnetic field is proportional to the current: $M \propto I$. Owing to this, we can measure the current from the angle of rotation of the coil if the scale is graduated appropriately. For this we must establish which angles of rotation of the pointer correspond to known values of the current.

?
1. Which forces are known as magnetic forces?
2. List the main properties of the magnetic field.
3. How do a closed loop and magnetic needle move in a uniform magnetic field?
4. Indicate the method of determining the direction of the magnetic induction.
5. How can the magnitude of the magnetic induction be determined?
6. Give the definition of magnetic field lines.
7. Which fields are known as vortex fields?
8. What is magnetic flux?
9. What is the operating principle of measuring instruments of the magnetoelectric type?

11.6. Ampère's Law

Magnetic field acts on all segments of a current-carrying conductor. The law determining the force acting on an individual segment of a conductor (current element) was established in 1820 by A. Ampère. Since a current element cannot be isolated, Ampère's experiments were carried out with closed conductors. By varying the shape of conductors and their arrangement, Ampère succeeded in deriving the expression for the force acting on an individual element of current.

MAGNITUDE OF AMPÈRE'S FORCE. Figure 224 shows a set-up which can be used to find out which factors determine the force acting on a current-carrying conductor.
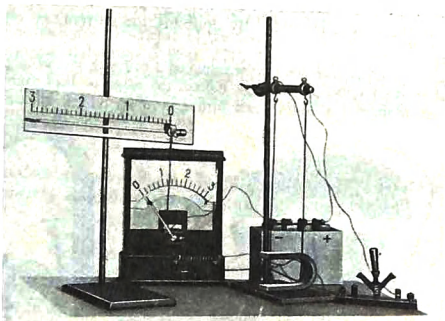

Fig. 224

The magnetic field is produced by a permanent magnet and mainly acts on a horizontal conductor. The force is measured with the help of a special balance connected to the conductor by two rods.

Doubling the current, we see that the force acting on the conductor is also doubled. Using one more magnet, we almost double the length of the segment of the conductor on which the magnetic field acts. The force is also doubled in this case. Using different magnets, we can establish that the magnetic (Ampère) force is proportional to the magnitude of vector $\vec{B}$.

Finally, Ampère's force depends on the angle between vector $\vec{B}$ and the conductor. This can be verified by varying the slope of the support on which the magnets rest so that the angle between the conductor and magnetic field lines changes.

Naturally, all the conclusions drawn above are of a qualitative nature. Let us now formulate the law correctly.

Suppose that the magnetic induction $\vec{B}$ forms an angle α with the direction of a current-carrying conductor (current element) (Fig. 225). (For the direction of the conductor, we take the direction of current in it.) Experiments show that the magnetic field, whose induction vector is directed along the current-carrying conductor does not produce any action on the current. Therefore, the magnitude of force depends only on that component of vector $\vec{B}$, which is perpendicular to the conductor, i.e. on $B_\perp = B \sin \alpha$, and does not depend on the component of vector $\vec{B}$, which is directed along the conductor.

The expression for the magnitude F of the force exerted on a small segment Δl of a conductor carrying a current I, by a magnetic field of induction $\vec{B}$ forming an angle α with the current element has the form

$$F = B \, | \, I \, | \, \Delta l \sin \alpha. \tag{11.4}$$

This expression is known as AMPÈRE'S LAW.

DIRECTION OF AMPÈRE'S FORCE. In the experiments considered above, vector $\vec{F}$ is normal to the current element and to vector $\vec{B}$. Its

Fig. 225   Fig. 226

direction is determined by the left-hand rule:

Place the left hand so that the component of vector B perpendicular to the conductor enters the palm and direct four fingers currentwise. Then the thumb at $90°$ to the fingers will indicate the direction of the force acting on the conductor (Fig. 226).

APPLICATION OF AMPÈRE'S LAW. Knowing the direction and magnitude of the force acting on any segment of a conductor, we can calculate the force acting on the entire closed conductor. For this purpose, we must find the sum of forces acting on all segments of the conductor. We can be completely sure that Ampère's law is valid because the magnitude of the resultant force calculated in this way for any closed conductor coincides with the experimentally determined value of this force.

Ampère's law is used for calculating the forces acting on current-carrying conductors in many appliances, in particular, in electric motors.

The operation of all electric motors is based on the utilization of Ampère's forces. Electric current flows in the winding of the rotating part of a motor (called the rotor). A powerful electromagnet produces a magnetic field which acts on current-carrying conductors and makes them move. Rotors are manufactured from steel plates, and the pole pieces of the electromagnet have such a shape that they concentrate the magnetic induction in the regions where the rotor winding is arranged.

Special systems ensure such a direction of current in the windings that the magnetic interaction creates a torque resulting in the continuous rotation of the rotor.

11.7. Action of Magnetic Field on a Moving Charge. Lorents Force

An electric current is a collective motion of charged particles.

Therefore, the action of a magnetic field on a current-carrying conductor is the result of action of the field on moving charged particles inside the conductor.

The force exerted by a magnetic field on a moving charged particle is called the Lorentz force, after the great Dutch physicist Lorentz, the founder of the electron theory of structure of matter. This force can be determined with the help of Ampère's law. The magnitude of the Lorentz force is equal to the ratio of the magnitude F of the force acting on a segment of the conductor of length Δl to the number N of particles in ordered motion over
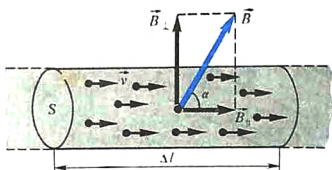
Fig. 227

this segment of the conductor:

$$F_L = \frac{F}{N}.$$

$$(11.5)$$

Let us consider a segment of a thin straight current-carrying conductor (Fig. 227). Suppose that the length Δl of the segment and its cross-sectional area S are so small that the magnetic induction $\vec{B}$ can be assumed to be constant within the limits of this segment. The current I in the conductor is related to the charge q_0 of a particle, the number density of charged particles (the number of charged particles per unit volume) n and the velocity v of their ordered motion through the following formula (see Sec. 9.1):

$$I = q_0 n v S. \qquad (11.6)$$

The magnitude of the force exerted by the magnetic field on the chosen current element is

$$F = |I| \Delta l B \sin \alpha.$$

Substituting into this formula expression (11.6) for current, we obtain

$$F = |q_0| nvS \Delta l B \sin \alpha = v|q_0| NB \sin \alpha,$$

where $N = nS \Delta l$ is the number of charged particles in the volume under consideration. Consequently, the magnetic field acts on each moving charge with the Lorentz force

$$F_L = \frac{F}{N} = |q_0| vB \sin \alpha, \qquad (11.7)$$

where α is the angle between the velocity vector and the magnetic induction. The *Lorentz force is at right angles to vectors* $\vec{B}$ *and* $\vec{v}$, its direction being determined by the same left-hand rule as the direction of Ampère's force.

Place the left hand so that the component of vector $\vec{B}$ perpendicular to the velocity of the charge enters the palm, direct four fingers along the motion of the positive charge (and opposite to it for the negative charge); then the thumb at a right angle to the four fingers points in the direction of the Lorentz force acting on the charge (Fig. 228).

An electric field acts on a charge q_0 with the force $\vec{F}_{el} = q_0 \vec{F}$. Thus, if the electric and magnetic fields are both present, the total force $\vec{F}$ acting on the
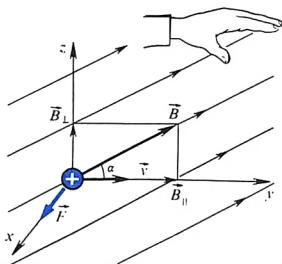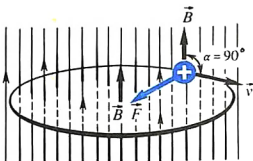
231

Fig. 228   Fig. 229

charge is given by[1]

$$\vec{F} = \vec{F}_{el} + \vec{F}_{L}.$$ (11.8)

Since the Lorentz force is at right angles to the velocity of a particle, it does no work. According to the theorem on kinetic energy (see "Senior Physics 1"), this means that the Lorentz force does not change the kinetic energy of the particle, and hence the magnitude of its velocity. Only the direction of the particle velocity changes under the action of the Lorentz force.

The action of the Lorentz force on moving electrons can be observed by bringing an electromagnet (or a permanent magnet) to a cathode-ray tube. By varying the current in the windings of the electromagnet, we notice that the deflection of the electron beam increases with the magnitude B of the magnetic induction of the field. If the direction of current in the electromagnet is reversed, the beam is deflected in the opposite direction.

The dependence of the Lorentz force on the angle α between the vectors $\vec{B}$ and $\vec{v}$ can be detected while observing the deflection of the electron beam with the variation of the angle between the axis of the magnet and the axis of the cathode-ray tube.

MOTION OF A CHARGED PARTICLE IN A UNIFORM MAGNETIC FIELD. Let us consider the motion of a particle having a charge q_0 in a uniform magnetic field $\vec{B}$ directed at right angles to the initial velocity $\vec{v}$ of the particle (Fig. 229). The Lorentz force depends on the magnitude of the velocity of the particle and the induction of the field. Since the magnetic field does not alter the magnitude of the velocity, the magnitude of the Lorentz force also remains unchanged. This force is perpendicular to the velocity, and hence determines the centripetal acceleration of the particle. The fact that the magnitude of the centripetal acceleration of a particle moving at a constant (in magnitude) velocity indicates that the particle moves uniformly in a circle of radius r. Let us determine this radius.

---

[1] The total force (11.8) exerted on a charged particle by an electromagnetic field is often called the Lorentz force. In this case, force (11.7) is called the magnetic component of the Lorentz force.
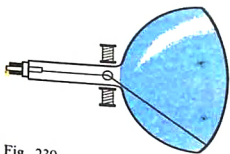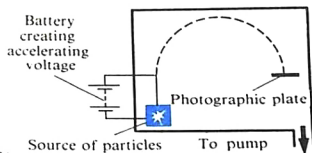
Fig. 230

Battery creating accelerating voltage

Photographic plate

Source of particles   To pump

Fig. 231

According to Newton's second law (Fig. 229), we have

$$mv^2 / r = |q_0| vB.$$

Hence

$$r = mv / |q_0| B. \qquad (11.9)$$

APPLICATION OF THE LORENTZ FORCE. The action of a magnetic field on a moving charge is widely used in modern engineering. It is sufficient to mention picture tubes (kinescopes) in which electrons flying to the screen are deflected by the magnetic field by special coils (Fig. 230).

Another application of the magnetic field lies in the separation of charges according to their charge-to-mass ratio. The obtained results make it possible to determine the mass of a particle precisely. The instruments whose operation is based on this principle are known as MASS SPECTROGRAPHS.

Figure 231 shows a schematic diagram of a simple mass spectrograph. The vacuum chamber of the instrument is placed in a magnetic field (the magnetic induction B is normal to the plane of the figure). Charged particles (electrons or ions) accelerated by the magnetic field move in an arc of a circle and get to a photographic plate, where they leave traces from which r can be measured to a high degree of accuracy. Knowing the radius of the trajectory, we can easily determine the charge-to-mass ratio of the ion. If the charge of the ion is known, it is not difficult to find its mass.

?

1. What is the magnitude of Ampère's force equal to?
2. Formulate the rule for determining the direction of Ampère's force.
3. Formulate the rule for determining the direction of the Lorentz force.
4. Write the formula for the magnitude of the Lorentz force.
5. Why does the Lorentz force leave the magnitude of the velocity of a charged particle unchanged?
6. What is the motion of a charged particle in a uniform magnetic field if its initial velocity is perpendicular to the magnetic field lines?

11.8. Magnetic Properties of Substances

PERMEABILITY. It is well known that magnetic fields are produced not only by electric currents but also by permanent magnets. Permanent magnets can be manufactured only from a few materials. However, all substances introduced into a magnetic field are
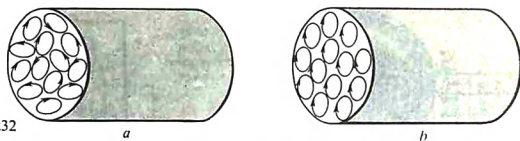
233

Fig. 232   *a*   *b*

magnetized, i.e. create a magnetic field themselves. As a result, the magnetic induction $\vec{B}$ in a homogeneous medium differs from the vector $\vec{B}_0$ at the same point of space in vacuum.

The ratio $B/B_0 = \mu$ characterizing the magnetic properties of a medium is known as the PERMEABILITY of the medium.

Thus, the magnetic induction in a homogeneous medium is

$$\vec{B} = \mu\vec{B}_0 \tag{11.10}$$

where μ is the permeability of the given medium[1].

AMPÈRE'S HYPOTHESIS. The factor responsible for the magnetic properties of substances was determined for the first time by the French scientist Ampère. Impressed by the effect observed in Oersted's experiments, viz. the rotation of a magnetic needle near a current-carrying conductor, Ampère proposed that the magnetism of the Earth is caused by currents flowing inside the globe. The most important conclusion was made: the magnetic properties of a body can be explained by the currents circulating in it. Later, Ampère arrived at the general conclusion: the magnetic properties of a body are determined by closed electric currents inside it. This decisive transition from the possibility of explaining the magnetic properties of bodies by currents to the rigorous statement that magnetic interactions are the interactions of currents, reflects the Ampère's boldness of scientific analysis. According to Ampère's hypothesis, elementary electric currents circulate within atoms and molecules. (It is well known now that these currents are due to the motion of electrons in atoms.) If the planes in which these currents circulate are arranged chaotically relative to one another as a result of random thermal motion of molecules (Fig. 232a), their actions are mutually balanced, and the body does not exhibit any magnetic properties. In the magnetized state, the elementary currents in a body are oriented so that their actions are added (Fig. 232b).

Ampère's hypothesis explains why a magnetic needle and a current loop behave identically in a magnetic field (see Sec. 11.2). The needle can be treated as an aggregate of small current loops having a similar orientation.

The bodies having a high permeability ($\mu \gg 1$) are called FERROMAGNETICS (iron, cobalt, nickel, rare-earth metals and many alloys). In these materials,

---

[1] Formula (11.10) is valid only for a homogeneous medium filling the entire space or for a special symmetry of a body, e.g., a homogeneous rod in a solenoid. When a body of an arbitrary shape is introduced in a magnetic field of induction $\vec{B}_0$, the induction in the substance will not be determined by formula (11.10). The dependence between $\vec{B}$ and $\vec{B}_0$ is much more complicated and is determined by the shape of the body and its orientation relative to $\vec{B}_0$.

the magnetic field is produced, however, not as a result of rotation of electrons about nuclei, but due to their "intrinsic rotation". An electron as if continuously moves around its axis, and since it bears a charge, it creates a magnetic field in addition to the field due to the orbital rotation.[1]

CURIE TEMPERATURE. The ferromagnetic properties of a ferromagnetic vanish at temperatures exceeding a certain temperature characteristic of the given ferromagnetic. This temperature is called the CURIE TEMPERATURE, or Curie point, after the French scientist who discovered this phenomenon. If we heat a magnetized nail to a high temperature, it loses its ability to attract ferrous materials. The Curie point is 753 °C for iron, 365 °C for nickel, and 1000 °C for cobalt. For some ferromagnetic alloys, the Curie temperature is less than 100 °C.

The first detailed investigations of the magnetic properties of ferromagnetics were carried out by the outstanding Russian physicist A. G. Stoletov (1839-1896).

FERROMAGNETICS AND THEIR APPLICATION. Although there are just a few ferromagnetics in nature, they are of great practical importance. Inserting an iron or steel core into a coil, we can enhance many times the magnetic field produced by it. The cores of transformers, generators, electric motors, etc. are made of ferromagnetics.

The permeability of ferromagnetic is not constant. It depends on the magnetic induction of the field.

When the external magnetic field is removed, a ferromagnetic remains magnetized, and is capable of creating a magnetic field in the space surrounding it. The ordered orientation of elementary currents does not vanish when the external field is switched off. This accounts for the existence of permanent magnets.

Permanent magnets are widely used in electrical measuring instruments, loudspeakers and telephones, sound-recording apparatus, magnetic compasses, and so on. (The application of permanent magnets was described in the course of "Junior Physics".)

Among ferromagnetic materials, there are so-called ferrites, i.e. substances which do not conduct an electric current. These are materials formed by ferric oxides combined with oxides of other elements. The first ferromagnetic material known to man (magnetic iron ore) is a ferrite.

?   1. Which quantity characterizes magnetic properties of a medium?
              2. What is the essence of Ampère's hypothesis?
              3. Which bodies are known as ferromagnetics?
              4. Where are ferromagnetic materials employed?

---

[1] We have to say "as if rotates" since an electron does not resemble a small ball in its properties. Its motion is governed by the laws of quantum, and not classical, mechanics. The intrinsic angular momentum (moment of momentum) of the electron is called the spin.

Fig. 233   *a*

b

Problems with Solutions

While solving problems on the magnetic field of current, you must be able to determine the direction of Ampère's force for given directions of the current and magnetic induction (or the direction of current for given directions of Ampère's force and magnetic induction) with the help of the left-hand rule. You must also know how to determine the direction of Lorentz force. The direction of the magnetic induction of a field is determined with the help of the right-hand screw rule.

Problems involving Ampère's law and the expression for the Lorentz force should be solved in the same way as problems in mechanics. In this case, in addition to mechanical forces, the Ampère and Lorentz forces should be taken into consideration.

1. A straight current-carrying wire is placed over the poles of a U shaped magnet (Fig. 233a). The wire can move in any direction. Which direction will it take?

Solution. We find the directions of magnetic induction vectors in various points and apply the left-hand rule. The wire will turn in the horizontal plane and simultaneously move into the space between the poles of the magnet (Fig. 233b).

2. A rectangular current loop shown in Fig. 234 has the following dimensions: $a = 3$ cm, $b = 5$ cm. The current in the loop is $I = 10$ A. The magnetic induction $B = 0.1$ T forms an angle $\alpha = 30°$ with the normal to the loop. Determine the moment of force exerted by the magnetic field on the loop.

Solution. Figure 235 is the top view on the section of the loop by a horizontal plane. In accordance with the left-hand rule, a couple of forces $\vec{F}_1$ and $\vec{F}_2$ normal to vector $\vec{B}$ acts on the sides of length b of the loop. This couple produces a torque about the axis passing through the middle of the loop. According to the left-hand rule, the forces acting on the sides of length a just stretch the loop.
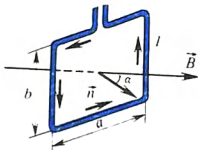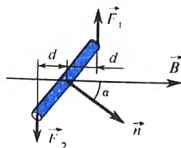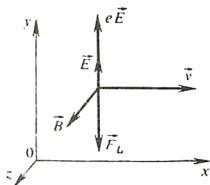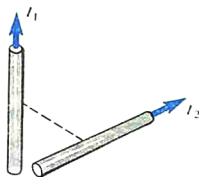

Fig. 234

Fig. 235

236

Fig. 236

Fig. 237

By Ampère's law, $F_1 = F_2 = F = |I|Bb$. The arm of each force is $d = (a/2)\sin\alpha$.

The moment of each force is $M_1 = F_1(a/2)\sin\alpha = (1/2)|I|BS\sin\alpha$ where $S = ab$ is the area of the loop. The total moment of force is $M = 2M_1 = |I|BS\sin\alpha$. For $\alpha = 90°$, the moment of force is maximum and coincides with the value of M given by formula (11.1) which was introduced as a definition of the magnitude of magnetic induction. Substituting the numerical values of the quantities, we obtain $M = 7.5 \times 10^{-4}$ N·m.

3. A proton moves along a straight path in the space where uniform and constant electric and magnetic fields exist simultaneously. The electric field strength is known to be $\vec{E}$. What is the induction $\vec{B}$ of the magnetic field?

Solution. The proton can move in a straight line in the following two cases:

(1) Vector $\vec{E}$ is directed along the proton path. Then vector $\vec{B}$ must also be directed along the path, and its magnitude can be arbitrary since the magnetic field will not act on the particle.

(2) Vectors $\vec{E}$, $\vec{B}$, and $\vec{v}$ are at right angles, and the force exerted on the proton by the electric field is equal and opposite to the Lorentz force exerted by the magnetic field on the proton (Fig. 236). Since $e\vec{E} + \vec{F}_L = 0$, $eE - evB = 0$ and $B = E/v$.

Exercise 14

1. Using the right-hand screw rule and the left-hand rule prove that unidirectional parallel currents attract each other while antiparallel currents repel each other.

2. Currents I_1 and I_2 are passed through two straight conductors crossed at right angles (Fig. 237). What will be the change in the relative arrangement of the conductors?

3. A current loop described in Problem 2 with solution is turned so that its normal forms an angle of $90°$ with the magnetic lines. Determine the moment of force acting on the loop.

4. A conductor having a length $l = 0.15$ m and carrying a current $I = 8$ A is normal to the magnetic induction of a uniform magnetic field whose magnitude $B = 0.4$ T. Find the work done in displacing the conductor by 0.025 m in the direction of Ampère's force.

5. Determine the direction of the magnetic induction in the .neck of a kinescope (see Fig. 230).

237

6. A particle having a charge q and mass m starts to move in a uniform magnetic field of induction $\vec{B}$ at a velocity $\vec{v}$. The initial velocity of the particle forms an angle α with vector $\vec{B}$. Show that the trajectory of the particle is a helix. What is the radius of this helix?

Main Points of Chapter 11

The interaction between electric currents, called the magnetic interaction, is realized through magnetic field. The basic characteristic of a magnetic field is the magnetic induction $\vec{B}$.

For the direction of the magnetic induction, we take the direction of the normal to a current loop which can be freely oriented in the magnetic field. This direction coincides with the direction of a magnetic needle in the field, i.e. with the direction of the line connecting the south and north poles of the needle.

The magnitude of magnetic induction is determined as the ratio of the maximum moment of force acting on a current loop in the magnetic field to the product of the current in the loop and its area.

The magnetic field lines envelop current-carrying conductors and are always closed. Fields having closed lines of force are known as vortex fields.

According to Ampère's law, a magnetic field exerts on a segment of a current-carrying conductor of length Δl a force whose magnitude is given by

$$F = B\,|\,I\,|\,\Delta l \sin \alpha,$$

where α is the angle between the segment of the conductor and the vector $\vec{B}$. The direction of the force is determined by the left-hand rule.

A moving particle experiences in a magnetic field the action of the Lorentz force whose magnitude is

$$F_L = |\,q_0\,|\,vB \sin \alpha,$$

where α is the angle between the velocity of the particle and the magnetic induction. The Lorentz force is perpendicular to the velocity of the particle and for this reason does no work.

All bodies are magnetized in a magnetic field, i.e. produce a magnetic field themselves. The ratio of the magnitude of magnetic induction in a homogeneous medium to the magnetic induction in vacuum is called the permeability of the medium and characterizes the magnetic properties of the material:

$$\mu = B/B_0.$$

In most substances, magnetic properties are pronounced weakly. Only ferromagnetic materials, which include iron, have a very high permeability ($\mu \gg 1$) which depends on the magnetic induction. Although the number of ferromagnetics is not very large, they are of great practical importance since they permit to increase magnetic induction hundreds of times without energy expenditures.

12 ELECTROMAGNETIC INDUCTION

12.1. Discovery of Electromagnetic Induction

Till now, we have considered electric and magnetic fields which do not vary with time. It was found that the electric field is produced by electric charges, while the magnetic field is generated by moving charges, i.e. by an electric current. Let us now investigate electric and magnetic fields varying with time.

The most important fact which has been revealed is a close relationship between electric and magnetic fields. *A time-varying magnetic filed generates an electric field, while varying electric fields generate magnetic fields.* Without this relationship, the electromagnetic forces would not be so diversified as they are in actual practice. There would be neither radio waves no light.

It was not by coincidence that a decisive step in the discovery of new properties of electromagnetic interactions was taken by Faraday, the founder of the theory of electromagnetic field. He was convinced that electric and magnetic phenomena have the same origin. Due to this, he made the discovery that later formed the basis of operation of generators in all electric power plants of the world, which convert the mechanical energy into the energy of electric current.

An electric current, Faraday argued, is capable of magnetizing a piece of iron. But can a magnet, in turn, cause an electric current?

For a long time, this relation could not be observed. The most difficult aspect was the apprehension of the main idea: only a moving magnet or time-varying magnetic field may excite an electric current in a coil.

Electromagnetic induction consists in the appearance of an electric current in a conducting loop which is either at rest in a time-varying magnetic field or moves in a constant magnetic field in such a way that the number of magnetic lines piercing the loop changes. This phenomenon was discovered on August 29, 1831. It is a rare case that the date of a remarkable discovery is known exactly.

The electromagnetic induction was first observed in two conductors, stationary relative to each other, during connection and disconnection of a circuit. Faraday, who believed that the relative motion of current-carrying
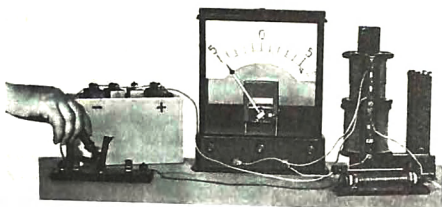
239

Fig. 238

conductors should lead to the same result as the connection or disconnection of the circuit, proved experimentally that a current appears when coils move relative to each other. Conversant with Ampère's idea, Faraday treated a magnet as an aggregate of small currents circulating in the molecules. On the 17th October, as was registered in his lab diary, an induction current was observed in a coil into which a magnet was being inserted (or withdrawn). During a month, Faraday discovered experimentally all significant features of electromagnetic induction.

At present, every student can repeat Faraday's experiments. For this we only need two coils, a magnet, a cell battery and a sufficiently sensitive galvanometer.

In the set-up shown in Fig. 238, a current is induced in one of the coils when the electric circuit of the other coil, which is at rest relative to the first one, is connected or disconnected. In the set-up shown in Fig. 239, the current in one of the coils is varied with the help of a rheostat. In Fig. 240a, current is induced when coils are moved relative to each other while in Fig. 240b, when a permanent magnet is moved relative to a coil.

Faraday was the first to note the general regularities of appearance of induced current in experiments which seem to be different.
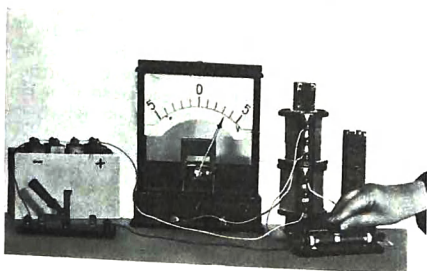

Fig. 239

Fig. 240

b
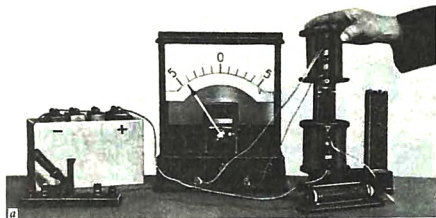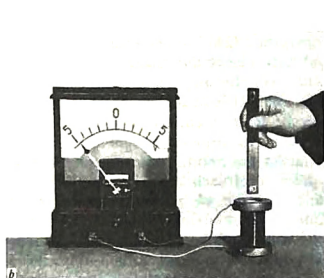
Fig. 241

A current is induced in a closed conducting loop when the number of magnetic field lines piercing the area bounded by the loop changes. The higher the rate of variation of the number of magnetic field lines, the stronger the induced current. The reason behind the change in the number of magnetic field lines is immaterial in this case. The number of magnetic field lines piercing the area bounded by a stationary loop may change as a result of a variation of current in a neighbouring coil (Fig. 238) or when a loop moves in a nonuniform magnetic field in which the density of lines of force changes in space (Fig. 241).

12.2. Direction of Induced Current. Lenz's Rule

Let us now determine the direction of induced current.
Connecting a coil in which a current is induced to a galvanometer, we note that the direction of this current depends on whether a magnet approaches the coil (e. g., with its north pole) or moves away from it (Fig. 240b).

The induced current of a certain direction interacts with the magnet. The current-carrying coil is similar to a magnet with north and south poles. The

Fig. 242

Fig. 243

direction of the induced current determines the end of the coil which plays the role of the north pole (the magnetic field lines emerge from it) and the end which becomes the south pole (magnetic field lines enter it). On the basis of the law 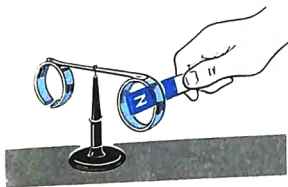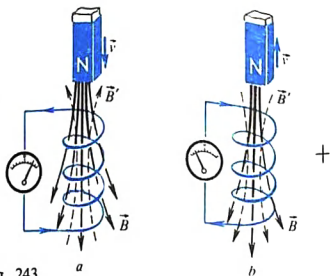of energy conservation, we can predict the case when the coil will attract the magnet and when it will repel the magnet.

INTERACTION OF INDUCED CURRENT WITH A MAGNET. If a magnet is brought closer to a coil, the induced current will certainly repel the magnet. A positive work must be done to bring the magnet closer to the coil. The coil behaves as a magnet facing the approaching magnet with the like pole. As is well known, like poles repel each other.

Let us suppose that the opposite is true, i.e. we push the magnet to the coil, and it is pulled inside it. This would be a violation of the energy conservation law. Indeed, the kinetic energy of the magnet would increase, and at the same time a current would be induced, which requires an expenditure of energy. The kinetic energy of the magnet and the energy of current would appear from without, involving no expenditure of energy.

On the contrary, when a magnet is moved away, the energy conservation law necessitates the appearance of an attractive force.

The validity of this conclusion can be verified with the help of an experiment represented in Fig. 242. Two conducting aluminium rings are fixed at the ends of a rod which can freely rotate about a vertical axis. One of the rings has a cut. If we bring a magnet to the ring without a cut, the current induced in it will have such a direction that the ring will be repelled from the magnet, and the rod will turn. If the magnet is moved away from the ring, it will be, on the contrary, attracted to the magnet. The magnet does not interact with the open ring since no current can be induced in it. The attraction or repulsion of a magnet by a coil is determined by the direction of current induced in it. Thus, the energy conservation law allows us to formulate the rule determining the direction of induced current.

What is the difference in the two experiments, the one in which the magnet approaches the coil, and the other in which the magnet is withdrawn? In the former case, the number of magnetic field lines piercing the turns of the coil (or, which is the same, the magnetic flux) increases (Fig. 243a), while in the latter case it decreases (Fig. 243b). In the former case, the magnetic field lines of

the magnetic field of the current induced in the coil emerge from the upper end of the coil so that it repels the magnet, while in the latter case, conversely, the magnetic field lines enter this end of the coil. These magnetic field lines are shown by the dashed lines in Fig. 243.

LENZ'S RULE. Thus, we come to the main point: as the magnetic flux through the turns of a coil increases, the induced current has such a direction that the magnetic field produced by it prevents the magnetic flux through the coil turns from increasing. Indeed, the magnetic induction $\vec{B}'$ of this field is directed against the magnetic induction $\vec{B}$ generating the induced current. On the other hand, if the magnetic flux through the coil decreases, the induced current produces a magnetic field of induction $\vec{B}'$, which enhances the magnetic flux through the turns of the coil.

This is the essence of the general rule for determining the direction of induced current which can be applied in all cases. This rule was established by the Russian physicist Lenz.

According to Lenz's rule, the current induced in a closed loop opposes the change in the magnetic field which causes it.

To determine the direction of induced current I_i in a loop with the help of Lenz's rule, one should proceed as follows.

(1) Determine the direction of the lines of magnetic induction $\vec{B}$ of the external field.

(2) Find out whether the magnetic flux of this field through the area bounded by the loop increases ($\Delta\Phi > 0$) or decreases ($\Delta\Phi < 0$).

(3) Determine the direction of lines of magnetic induction $\vec{B}'$ of the magnetic field produced by the induced current I_i. According to Lenz's rule, these lines must be directed against the lines of $\vec{B}$ for $\Delta\Phi > 0$ and must be unidirectional with them for $\Delta\Phi < 0$.

(4) From the direction of the lines of magnetic induction $\vec{B}'$, determine the direction of the induced current I_i with the help of the right-hand screw rule.

Try to use this procedure for determining the direction of induced currents in the examples considered above.

12.3. Faraday's Law of Electromagnetic Induction

Let us now formulate the law of electromagnetic induction quantitatively. Faraday's experiments have revealed that the current I_i induced in a conducting loop is proportional to the rate of variation of the number of lines of magnetic induction $\vec{B}$ piercing the area bounded by the loop. This statement can be formulated more accurately by using the concept of magnetic flux.

Magnetic flux was visually interpreted as the number of magnetic field lines piercing a surface of area S (see Sec. 11.4). Therefore, the rate of variation of this number is precisely the rate of variation of the magnetic flux.

If during a small time interval Δt the magnetic flux changes by $\Delta\Phi$, the rate of its variation is $\Delta\Phi/\Delta t$. Consequently, the statement that directly follows from experiments can be formulated as follows: the induced current

is proportional to the rate of variation of the magnetic flux through the surface bounded by the loop:

$$I_i \propto \Delta\Phi/\Delta t. \tag{12.1}$$

INDUCED EMF. An electric current is known to appear in a circuit when extraneous forces act on free charges in a conductor. The work done by these forces in carrying a unit positive charge along a closed loop is known as electromotive force. Therefore, when the magnetic flux through the surface bounded by the loop changes, extraneous forces emerge, whose action is characterized by an emf called the induced emf. It is denoted by $\mathscr{E}_i$.

According to Ohm's law for a closed circuit, $I_i = \mathscr{E}_i/R$. The resistance of a conductor does not depend on the magnetic flux variation. This means that expression (12.1) is valid just because the induced emf is proportional to $\Delta\Phi/\Delta t$.

FARADAY'S LAW OF ELECTROMAGNETIC INDUCTION. This law is formulated not for current but for induced emf. Such a formulation of the law expresses the essence of the phenomenon of electromagnetic induction which does not depend on the properties of conductors in which current is induced. According to the law of electromagnetic induction, the emf induced in a closed loop is equal in magnitude to the rate of variation of the magnetic flux through the surface bounded by the loop:[1]

$$|\mathscr{E}_i| = |\Delta\Phi/\Delta t|.$$

How can we take into account in this law the direction of induced current (or the sign of the induced emf) determined by Lenz's rule?

Figure 244 shows a closed loop. We shall assume that the anticlockwise circumvention of the loop has the positive direction. The normal $\vec{n}$ to the loop forms a right-hand screw with the direction of circumvention.

Suppose that the magnetic induction $\vec{B}$ is directed along the normal to the loop and increases with time. Then $\Phi > 0$ and $\Delta\Phi/\Delta t > 0$.

---

[1] In experiments, only the linear dependence of $\mathscr{E}_i$ on $\Delta\Phi/\Delta t$ is proved. However, the proportionality factor can be taken equal to unity since the unit for measuring the magnetic flux has not been introduced so far.
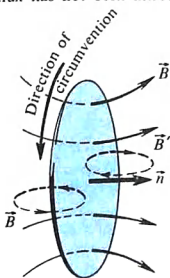

Fig. 244
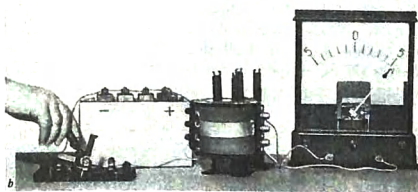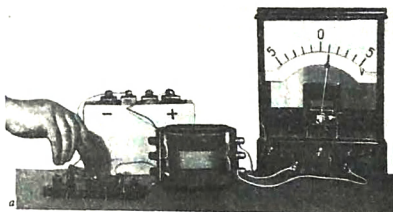
Fig. 245

According to Lenz's rule, the induced current produces a magnetic flux $\Phi' < 0$. The magnetic field lines of the induced current are shown in Fig. 244 by dashed lines. Consequently, in accordance with the right-hand screw rule, the induced current I_i is directed clockwise (i.e. against the positive direction of circumvention), and the induced emf is negative. Therefore, in the expression of the law of electromagnetic induction there should be the minus sign indicating that $\mathscr{E}_i$ and $\Delta\Phi/\Delta t$ have different signs:

$$\mathscr{E}_i = -\Delta\Phi/\Delta t. \tag{12.2}$$

MEASUREMENT OF PERMEABILITY OF IRON. Electromagnetic induction can be used for measuring the permeability of iron and other ferromagnetics.

If we insert an iron core into a long coil, then according to formula (11.10), the magnetic induction will increase μ times. Consequently, the magnetic flux and induced emf will increase in the same proportion.

At the moment of disconnection of a circuit feeding a magnetizing coil by a direct current, a current registered by a galvanometer is induced in another

coil wound around the first one (Fig. 245a). If an iron core is inserted into the coil, the deflection of the pointer of the galvanometer will be μ times larger (Fig. 245b). Measurements show that the magnetic flux can be increased thousands of times by introducing an iron core into the coil. This means that the permeability of iron is very large.

? 　　　　 1. What is the essence of electromagnetic induction?
　　　　　 2. How is the direction of induced current determined?
　　　　　 3. Formulate the law of electromagnetic induction.

12.4. Vortex Electric Field

An emf is induced either in a stationary conductor placed in a time-varying magnetic field or in a conductor moving in a magnetic field which does not vary with time. In both cases, the value of the emf is determined by the law (12.2), but the origin of the emf is not the same. Let us first consider the former case.

Let us take a transformer, viz. two coils wound around a core. Connecting the primary coil to an a.c. circuit, we obtain a current in the secondary coil (Fig. 246), if the latter is closed. The electrons in the secondary winding start to move. Which forces make them move? The magnetic field piercing the coil cannot do it by itself since the magnetic field acts only on moving charges (in this respect it differs from the electric field), and the conductor containing electrons is stationary[1].

In addition to the magnetic field, an electric field also acts on the charges. Unlike the magnetic field, electric field can act on stationary charges as well. But the field considered above (electrostatic and stationary field) is produced by electric charges, while the current is induced by a varying magnetic field. This leads to the assumption that in a stationary conductor, electrons are set in motion by an electric field, and this field is generated directly by a varying magnetic field. Thus, a new fundamental property of field is established: a magnetic field varying with time generates an electric field. This conclusion was drawn for the first time by Maxwell.

The phenomenon of electromagnetic induction is now seen by us from a new viewpoint. The essence of this phenomenon is the generation of an electric field by a magnetic field. Hence the presence of a conducting loop, say, a coil, does not alter the crux of the matter. A conductor with a store of free electrons (or other particles) only makes it possible to detect the emerging electric field. The field sets in motion the electrons in the conductor, and thus manifests itself. Essentially, electromagnetic induction in a stationary conductor consists in the emergence of an electric field which makes electrons move, rather than in the appearance of induced current.

---

[1] Actually, the matter is more complicated. Electrons are in random thermal motion even in a stationary conductor. However, the mean velocity of this motion is zero. Accordingly, the current induced by the magnetic field must also be zero.
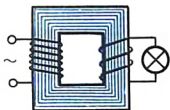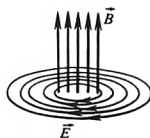
Fig. 246

Fig. 247

The electric field generated by a variation of a magnetic field has a completely different structure in comparison with an electrostatic field. It is not related directly to electric charges, and its lines of force cannot start or terminate on them. In general, these lines have neither beginning nor end and are closed like the magnetic field lines. This is the so-called VORTEX ELECTRIC FIELD. (Fig. 247).

The direction of lines of a vortex electric field coincides with the direction of induced current. The force exerted by a vortex electric field on a charge q is, as before, $\vec{F} = q\vec{E}$. But unlike the stationary electric field, the work of the vortex field on a closed path differs from zero. Indeed, when a charge is moved along a closed line of force of the vortex electric field (Fig. 247), the work has the same sign on all segments since the force and the displacement coincide in direction. The work of a vortex electric field in the displacement of a unit positive charge along a closed stationary conductor is numerically equal to the emf induced in this conductor.

INDUCED CURRENTS IN BULK CONDUCTORS. Induced currents reach especially high values in bulk conductors having a small resistance.

These currents, known as FOUCAULT (or eddy) CURRENTS after the French scientist who investigated them, can be used for heating conductors. The construction of induction furnaces is based on this principle. Such furnaces are widely employed for melting metals in vacuum.

However, in many engineering devices eddy currents lead to energy losses associated with the liberation of heat. For this reason, iron cores of transformers, electric motors and generators are made of individual insulated plates. The surfaces of the plates must be normal to the direction of vortex electric field. In this case, the resistance of the plates to electric current will be maximum.

APPLICATION OF FERRITES. It was mentioned in Sec. 11.8 that there exist magnetic insulators, viz. ferrites. Eddy currents do not appear in them upon a reversal of magnetization. As a result, energy losses associated with heat evolution are reduced to a minimum. This is of special importance for radio engineering devices operating in the range of very high frequencies (of the order of 10^6 Hz). In this case, the use of cores assembled from individual plates does not produce the required effect since large Foucault currents are generated in every plate. For this reason, ferrites are used for manufacturing cores for high-frequency transformers, magnetic aerials of transistors, and so on. Ferrite cores are obtained from a mixture of powdered ingredients. The mixture is pressed and subjected to thermal treatment.

Besides, it is essential that a rapid variation of magnetic field induces

247

a current in ordinary ferromagnetics. In accordance with Lenz's rule the magnetic field of this current opposes the variation of the magnetic flux in the core of the coil. As a result, the magnetic flux practically remains unchanged, and no magnetization reversal is observed in the core.

Eddy currents do not appear in ferrites and hence their magnetization can be easily reversed. This property of ferrites is used in many computers for creating memory of temporary storage (i.e. for rapidly recording and erasing information).

12.5. Induced EMF in Moving Conductors

If a conductor moves in a constant magnetic field, an emf is induced in it not due to a vortex field but for some other reason.

During the motion of the conductor, its free electrons move together with it. Therefore, the magnetic field acts on the charges with the Lorentz force. It is this force that causes the displacement of charges within the conductor. Consequently, in this case the induced emf is of the "magnetic origin".

Lorentz force causes the displacement of electrons in moving conductors at many electric power plants in the world.

Let us calculate the emf induced in a rectangular loop placed in a uniform magnetic field (Fig. 248). Let the side $|MN| = l$ of the loop slide at a constant velocity $\vec{v}$ along the sides NC and MD, remaining all the time parallel to the side CD. The magnetic induction $\vec{B}$ of the uniform field is normal to the conductor MN and forms an angle α with the direction of its velocity.

The magnitude of the force exerted by the magnetic field on a moving charged particle (see Sec. 11.7) is

$$F_L = |q_0| vB \sin \alpha. \tag{12.3}$$

This force is directed along MN. The work done by the Lorentz force over a distance l is positive and given by[1]

$$A = F_L l = |q_0| vBl \sin \alpha.$$

The electromotive force induced in the conductor MN is the ratio of the work on the displacement of a charge q_0 to this charge:

$$\mathscr{E} = \frac{A}{q_0} = vBl \sin \alpha. \tag{12.4}$$

This formula is valid for any conductor of length l, moving at velocity $\vec{v}$ in a uniform magnetic field.

Emf's induced in other conductors of the loop are equal to zero since they do not move. Consequently, the emf induced in the entire loop $MNCD$ is

---

[1] This is not the total work done by the Lorentz force. There also exists a component of this force antiparallel to the velocity $\vec{v}$ of the conductor. This force decelerates the conductor and does a negative work. As a result, the total work done by the Lorentz force is equal to zero.
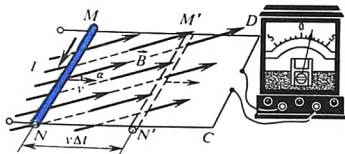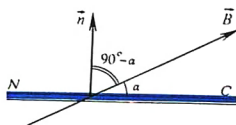
Fig. 248

Fig. 249

equal to $\mathscr{E}_i$ and remains constant if the velocity $\vec{v}$ does not change. The electric current in this case increases since the total resistance of the loop decreases as the conductor MN moves to the right.

On the other hand, the induced emf can be calculated with the help of the law of electromagnetic induction (12.2). Indeed, the magnetic flux through the loop $MNCD$ is

$$\Phi = BS \cos(90° - \alpha) = BS \sin \alpha,$$

where $90° - \alpha$ is the angle between the vector $\vec{B}$ and the normal $\vec{n}$ to the surface of the loop (Fig. 249), while S is the area of the loop $MNCD$. If we assume that at the initial instant of time ($t = 0$) the conductor MN is at a distance NC from the conductor CD (Fig. 248), the area S changes with time during the displacement of the conductor as follows:

$$S = |MN|(|NC| - vt).$$

Over a time Δt, the area bounded by the loop changes by $\Delta S = -lv\Delta t$. The minus sign indicates that it decreases. The change in the magnetic flux over this time is $\Delta \Phi = -Blv\Delta t \sin \alpha$. Consequently,

$$\mathscr{E}_i = -\frac{\Delta \Phi}{\Delta t} = Blv \sin \alpha.$$

If the loop $MNCD$ moves as a whole in a uniform magnetic field, retaining its orientation relative to the vector $\vec{B}$, the induced emf in the loop will be zero since the flux Φ through the area bounded by the loop does not change. This can be explained as follows. During the motion of the loop, forces (12.3) emerge in conductors MN and CD, which act on electrons in the directions from N to M and from C to D. The total work done by these forces in the clockwise (or anticlockwise) circumvention of the loop is zero.

?
1. What is the origin of the extraneous force responsible for inducing a current in a stationary conductor?
2. What is the difference between a vortex electric field and an electrostatic or stationary field?
3. What is the origin of the extraneous force causing an induced current in a moving conductor?
4. Which currents are known as Foucault currents?
5. What are the advantages of ferrites over ordinary ferromagnetics?

12.6. Self-induction. Inductance

SELF-INDUCTION. If an alternating current flows in a coil, the magnetic flux piercing the coil changes. Therefore, an emf is induced in the same conductor which carries the electric current. This phenomenon is known as SELF-INDUCTION. The role of the conducting loop in self-induction is dual: the current causing induction flows in it, and an emf is induced in the same loop. *A varying magnetic field induces an emf in the same conductor in which the current producing this field flows.*

According to Lenz's rule, as the current increases the vortex electric field strength is directed oppositely to the current. Hence, at this moment the vortex field prevents the current from increasing. On the contrary, at the moment when the current decreases, the vortex field sustains the current.

As a result, when a circuit containing a constant emf is closed, a definite value of the current is attained gradually and not instantaneously (Fig. 250). On the other hand, when a source of current is disconnected, the current in the circuit does not cease at once. The emf of self-induction emerging in the circuit may even exceed the emf of the current source since the current and its magnetic field vary very rapidly at the moment of disconnection.

Self-induction can be observed in simple experiments. Figure 251 represents a circuit containing two identical lamps connected in parallel. One of them is connected to a current source through a resistor of resistance R, while the other is series-connected to a coil L with an iron core.

When the key K is closed, the first lamp flares up almost instantaneously and the second one, with a noticeable delay. The emf of self-induction is large in the subcircuit containing this lamp, and the current cannot immediately reach its maximum value.

The emf of self-induction emerging during the disconnection of a circuit can be observed with the help of a circuit shown schematically in Fig. 252. When the key K is opened, an emf of self-induction appears in coil L, which sustains the initial current. As a result, at the moment of disconnection a current will flow through the galvanometer (dashed arrow), whose direction is opposite to that of the current existing before the disconnection (solid arrow). Current emerging during the disconnection of a circuit may exceed the current flowing through the galvanometer when the key K is closed. This means that $\mathscr{E}_{si} > \mathscr{E}$ of the battery.

THE ANALOGY BETWEEN SELF-INDUCTION AND INERTIA. Self-induction is similar to inertia encountered in mechanics. Indeed, due to inertia a body acquires a certain velocity under the action of force gradually and not instantaneously. The body cannot be stopped instantaneously
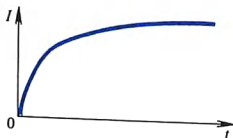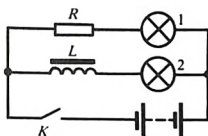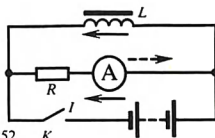

Fig. 250

Fig. 251   *K*

Fig. 252   *K*

irrespective of the magnitude of the braking force. Similarly, due to self-induction the current cannot attain its definite value instantaneously and increases gradually. By disconnecting a current source, we cannot terminate the current at once. Self-induction sustains it during a certain time in spite of the resistance of the circuit.

Further, according to the laws of mechanics, work must be done to increase the velocity of a body. During braking, the body itself does work. In the same way, to generate a current, a work against vortex electric field must be done, while during the attenuation of current this field does a positive work itself.

INDUCTANCE. The magnitude B of the magnetic induction of a field produced by the current is proportional to the current. Since the magnetic flux is proportional to B, we have $\Phi \propto B \propto I$.

Thus, we can state that

$$\Phi = LI, \tag{12.5}$$

where L is the proportionality factor between the current in a conducting loop and the magnetic flux piercing it. The quantity L is known as the INDUCTANCE OF THE LOOP, or the self-induction COEFFICIENT.

Using the law of electromagnetic induction and expression (12.5), we obtain the following equality:

$$\mathscr{E}_{si} = -\frac{\Delta\Phi}{\Delta t} = -L\frac{\Delta I}{\Delta t}, \tag{12.6}$$

if we assume that the shape of the loop remains unchanged and the magnetic flux changes only due to the change in the current.

It follows from formula (12.6) that inductance is a physical quantity numerically equal to the emf of self-induction appearing in the loop as a result of a change in the current by 1 A during 1 s.

Like capacitance, inductance depends on geometrical factors, viz. the shape and size of the conductor, and does not depend explicitly on the current in the conductor. Besides the conductor geometry, the inductance depends on the magnetic properties of the medium in which the conductor is placed.

The SI unit of inductance is a henry (H). A conductor has an inductance of 1 H if a change in the current in it by 1 A per second causes an emf of self-induction of 1 V:

$$1 \text{ H} = 1 \text{ V}/(1 \text{ A/s}) = 1 \text{ V}\cdot\text{s/A}.$$

Self-induction plays an important role in electrical and radio engineering.

17*

The inductance of a circuit considerably affects the passage of an alternating current in the circuit. This question will be treated in greater detail in the next volume of the book.

12.7. Energy of Magnetic Field Generated by a Current

Let us determine the energy of electric current in a conductor. In accordance with the energy conservation law, the energy of the current is equal to the energy that has to be spent by a current source (voltaic cell, generator at a power plant, etc.) to produce the current. When the current is discontinued, this energy is liberated in a certain form.

Let us find out why energy must be spent, i.e. work must be done to produce a current. This is explained by the fact that when the circuit is closed and the current increases, a vortex electric field is induced in the conductor, which is directed against the electric field existing in the conductor due to the current source.[1] For a current to reach the value I, the current source must do work against the vortex field forces. This work just increases the energy of the magnetic field produced by the current.

When the circuit is disconnected, the current attenuates, and the vortex field does a positive work. The energy stored by the current is liberated. This can be observed in the form of an intense spark appearing during a disconnection of a circuit with a large inductance (Fig. 253).

We can write the expression for current I flowing in a circuit of inductance L by analogy between inertia and self-inductance mentioned in the previous section.

While self-induction is similar to inertia, the inductance must play the same role in the process of generating current as the mass during an increase in velocity in mechanics. In electrodynamics, the part of the velocity of a body is played by current I, viz. the quantity characterizing the motion of electric charges.

---

[1] This electric field is produced by charged particles on the surface of the conductor, unlike the vortex field generated by a varying magnetic field.
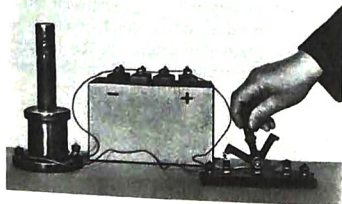

Fig. 253

Following this analogy, we can assume that the energy W_m of the magnetic field produced by a current is similar to the kinetic energy $mv^2/2$ of the body in mechanics:

$$W_m = LI^2/2. \tag{12.7}$$

This is the expression for the energy of the current, which can be obtained as a result of calculation.

The energy of current (12.7) is expressed in terms of the geometric characteristic L of the conductor and the current I in it. However, this energy can be expressed in terms of the field characteristics as well. Calculations show that the energy density of a magnetic field (i.e. the energy per unit volume) is proportional to the squared magnetic induction in the same way as the electric field energy density is proportional to the squared electric field strength.

?

1. What is self-induction?
2. What is the direction of the vortex electric field in a conductor relative to the direction of current in it when the current increases? decreases?
3. Give the definition of the inductance of a conductor.
4. What is the SI unit for inductance?
5. Write the expression for the emf of self-induction.
6. Why should the energy of a current source be spent to generate a current?
7. Write the expression for the energy of electric current.

Problems with Solutions

To solve the problems of this chapter, you must be able to determine the direction of induced current with the help of Lenz's rule. The induced emf is calculated with the help of the law of electromagnetic induction (12.2) or (for a conductor moving in a magnetic field) by formula (12.4). The emf of self-induction is calculated by formula (12.6), while the energy of current, is determined by formula (12.7). In some problems, Ampère's law should be used.

1. A rectangular loop $ABCD$ performs a translatory motion in the magnetic field produced by a straight long conductor (Fig. 254). Determine the direction of current induced in the loop if the latter moves away from the conductor. Which forces act on the loop?

Solution. The magnetic induction of the magnetic field produced by current I is normal to the plane of the loop and directed away from us. When the loop moves away from the current-carrying conductor, the magnetic flux through the area of $ABCD$ decreases ($\Delta\Phi < 0$). Consequently, the magnetic induction $\vec{B}'$ of the magnetic field produced by current I_i is directed, in accordance with Lenz's rule, away from us like the vector $\vec{B}$. With the help of the right-hand screw rule, we find that the current induced in the loop is directed clockwise.
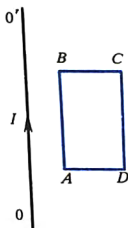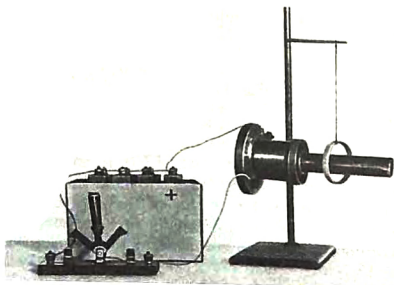
Fig. 254   Fig. 255

The interaction of the current in the loop with the straight current leads to the emergence of forces acting on the conductors of the loop. Applying the left-hand rule, we can see that, first, these forces stretch the loop, tending to increase its area and, second, create a resultant force directed towards the straight conductor. Both these actions "strive" to prevent the magnetic flux through the loop from decreasing.

2. A superconducting ring is placed in a uniform magnetic field whose induction increases from zero to B_0. The plane of the ring is perpendicular to the magnetic field lines. Determine the current induced in the ring, if its radius is r and inductance is L.

Solution. Since the resistance of the ring is zero, the total electromotive force in it must also be zero. Otherwise, according to Ohm's law, the current would be equal to infinity. Thus, the change in the magnetic flux of the external magnetic field is equal and opposite to the change in the magnetic flux produced by the induced current: $\Delta\Phi = L\Delta I$. Considering that the flux Φ_0 increases from 0 to $\pi r^2 B_0$, while the induced current varies from 0 to I, we obtain

$$\pi r^2 B_0 = LI, \quad \text{whence} \quad I = \frac{\pi r^2 B_0}{L}.$$

Exercise 15

1. In the circuit represented in Fig. 238, the key has been just closed. The current in the lower coil flows counterclockwise if viewed from the top. What is the direction of the current in the upper coil provided that it is fixed?
2. A magnet (Fig. 240b) is pulled out from a coil. Determine the direction of the current induced in the coil.
3. Determine the direction of the current induced in a closed ring at the moment a magnet is brought to it (Fig. 242).
4. The current in conductor OO' (Fig. 254) decreases. Determine the direction of the current induced in the fixed loop $ABCD$ and the direction of forces acting on each side of the loop.

254

5. A metallic ring can move freely along the core of a coil connected to a d.c. circuit (Fig. 255). What will be observed at the moment of connection and disconnection of the circuit?

6. The current in a coil increases in direct proportion to time. What is the nature of the time dependence of current in another coil which is inductively coupled with the first coil?

7. In which case do oscillations of the pointer of a magnetoelectric instrument attenuate quicker: when the instrument terminals are short-circuited or when they are disconnected?

8. The magnetic flux through the conducting loop having a resistance of 3×10^{-2} Ω has changed by 1.2×10^{-2} Wb over 2 s. What is the current in the conductor if the magnetic flux has been changing uniformly?

9. An aeroplane flies horizontally at a velocity of 900 km/h. Find the potential difference across the edges of its wings if the magnitude of the vertical component of the magnetic induction of the magnetic field of the Earth is 5×10^{-5} T and the wingspan of the aeroplane is 12 m.

10. A current of 4 A has been established in a coil having an inductance of 0.15 H and a small resistance r. A resistor of resistance $R \gg r$ is connected in parallel to the coil. What amount of heat is liberated in the coil and resistor after the current source has been rapidly disconnected?

Main Ponts of Chapter 12

Electromagnetic induction consists in the emergence of an induced emf in a closed loop upon a change in a magnetic flux through the area bounded by this loop. This phenomenon forms the basis of the operation of generators in all electric power plants in the world. According to Faraday's law of electromagnetic induction, the emf induced in a closed loop is equal to the rate of variation of the magnetic flux with the minus sign:

$$\mathcal{E} = -\Delta\Phi/\Delta t.$$

In accordance with Lenz's rule, the current induced in a closed loop opposes the change in the magnetic flux which causes this current.

In a stationary conductor, the extraneous force acting on charges is a vortex electric field generated by a varying magnetic field. In a moving conductor, the extraneous force is the Lorentz force acting on charged particles moving together with the conductor.

An important particular case of electromagnetic induction is self-induction. In this case, a varying magnetic field induces an emf in the conductor carrying the current which produces this field.

The emf of self-induction is proportional to the rate of variation of current in the conductor: $\mathcal{E}_{si} = -L(\Delta I/\Delta t)$.

The proportionality factor L is called the inductance. It depends on the size and shape of a conductor, as well as on the properties of the medium in which the conductor is placed. Inductance is measured in henrys: 1 H = 1 V·s/A.

The energy of the magnetic field produced by the current $W_m = LI^2/2$, is equal to the work that should be done by the source to generate a given current.

Conclusion

The study of electrodynamics is not completed by an acquaintance with electromagnetic induction. So far, we considered only electromagnetic phenomena for stationary electric charges (constant electric field) and charges moving at a constant velocity (constant magnetic field). Time-varying magnetic field was analyzed only briefly while dealing with the material in the chapter on electromagnetic induction. A very wide, most interesting and important for applications field of varying and fast-varying electromagnetic phenomena has not been considered yet.

In the next volume, we shall first deal with the properties, production and applications of alternating electric current on the basis of the fundamental laws of electrodynamics presented in this book. Then another fundamental property of electromagnetic field, viz. the generation of a magnetic field by a varying electric field, will be considered. This phenomenon is inverse to electromagnetic induction. Finally rapidly varying electromagnetic fields and electromagnetic waves will be investigated. This will conclude your acquaintance with electrodynamics, and a complete pattern of electromagnetic processes in nature will generally become clear to you.

Practicals

1. Experimental Verification of Boyle's Law

Instruments and materials: a 300-350 mm long glass tube having a diameter of 8-10 mm and sealed at one end, a cylindrical vessel (40-50 mm in diameter and 350-400 mm long) or a glass tube closed by a cork at the bottom, a ruler with a millimetre scale, a holder, a barometer (one instrument is enough for a group).

HINTS. 1. Fill the cylindrical vessel fixed by the holder with water.
2. Measure the atmospheric pressure with the help of the barometer.
3. Measure the volume of air in the glass tube (in conventional units corresponding to the length measured by the ruler).
4. Determine the product of the pressure of air by its volume.
5. Immerse the glass tube with its sealed end up into the water.
6. Measure the new volume of air in the tube.
7. Measure the difference in water levels in the vessel and the tube.
8. Calculate the new air pressure in the tube.
9. Calculate the product of the pressure of the air by its volume.
10. Repeat the experiment several times.
11. Tabulate the results of measurements.

No. of experiment	Pressure p	Volume V	pV

Compare the obtained results and verify Boyle's law.

2. Determination of the Surface Tension of a Liquid

While calculating the surface tension, it is convenient to make use of the relation between the force of surface tension and the length of the surface layer boundary: $F = \sigma L$.

Instruments and materials: a beam balance with weights, a holder, a test tube with sand, vernier calipers or a ruler with a millimetre scale, a sheet of paper, and a wire or wire frame suspended by threads (Fig. 256).

Fig. 256
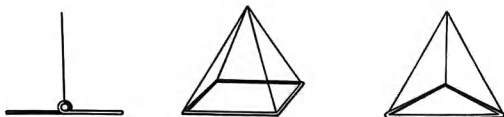
HINTS. 1. Fix the beam balance by the clutch of the holder.

2. Measure the length of the wire (perimeter of the frame) by the vernier calipers or scale ruler.

3. Attach the thread with the wire or frame to a pan of the balance and balance it with the help of sand (pour sand on a sheet of paper placed on the pan).

4. Set the wire (frame) in the horizontal position.

5. Put a glass of water under the pan so that the water surface is at 1-2 cm from the wire.

6. Lower slowly the wire or frame until it touches the surface of the water and "adheres" to it.

7. Add carefully some more sand until the wire or frame is "detached" from the water surface.

8. Dry the wire or frame with filter paper and balance the beam again with the help of weights.

9. While determining the mass m of the sand, the spread in the values may be caused by various factors (wire or frame is detached from the water surface not simultaneously at all points, the beam of the balance may oscillate, and so on). In such a case, several measurements should be made and the mean value $m_{av} = (m_1 + m_2 + m_3 + \ldots)/n$ should be determined.

In order to estimate the error of measurements, the mean absolute error is calculated:

$$\Delta m_{av} = \frac{|m_1 - m_{av}| + |m_2 - m_{av}| + |m_3 - m_{av}| + \ldots}{n}.$$

Repeat the experiment five times and tabulate the results of measurements:

n	m, kg	m_{av}, kg	$\Delta m = |m - m_{av}|$, kg	Δm_{av}, kg	l, m
1					
2					
3					
4					
5					

10. Calculate the surface tension

$$\sigma_{av} = \frac{m_{av} g}{2l}.$$

11. Since relative errors are added upon division or multiplication of measured quantities, the relative error in the measurement of the surface

258

tension coefficient can be found as follows:

$$\varepsilon_\sigma = \varepsilon_m + \varepsilon_g + \varepsilon_l.$$

While calculating ε_m, the following circumstance should be taken into account. The theory of measurements shows that when ten (or more) experiments have been carried out under invariable conditions, the mean absolute error can be used to estimate the error associated with a random spread in values in the process of measurement. If five experiments have been carried out, the mean absolute error should be increased three-fold for a correct estimation of the maximum error of measurements. In this work, $\varepsilon_m = 3\Delta m_{av}/m_{av}$ (the errors introduced by the weights and balance can be neglected).

If the value of g is rounded to 9.8 m/s^2, then $\varepsilon_g \simeq 0.002$ can be neglected in calculations.

While calculating $\varepsilon_l = \Delta l/l$, we can assume that Δl does not exceed 1 mm (if the ruler with a millimetre scale is being used). If ε_l is considerably less than ε_m, we can neglect ε_l as well.

12. Calculate $\Delta\sigma = \sigma_{av} \cdot \varepsilon_\sigma$ and write the result in the form

$$\sigma_{av} - \Delta\sigma \leqslant \sigma \leqslant \sigma_{av} + \Delta\sigma.$$

3. Determination of the Modulus of Elasticity of Rubber

Instruments and materials: a holder, the pan of a laboratory-type tribometer, weights, vernier caliper or slide gauge, a ruler with a millimetre scale, two 40-cm long rubber cords with different cross sections.

HINTS. 1. Suspend one end of a rubber cord in a holder.

2. Attach the pan for weights at the lower end of this cord.

3. Draw with a pencil or pen two thin marks at about 20 cm from each other in the middle of the cord.

4. Measure the initial distance l_0 between the marks.

5. Measure the diameter of the cord between the marks (without compressing rubber) and calculate the area S of its cross section.

6. Load the pan with weights, write down their weight F and measure the distance l between the marks after the extension of the cord has been completed.

7. Repeat the experiment, gradually increasing the load to the maximum value and then reducing it. Determine the absolute elongation $\Delta l = l - l_0$ of rubber for each loading.

8. Tabulate the results of measurements:

Cross-sectional area of cord, S, m^2	Deforming force (load) F, N	Length of cord (between marks), m		Absolute elongation l of cord (between marks), m	
		with increasing load	with decreasing load	with increasing load	with decreasing load

259

9. Using the obtained results, plot the graph for the extension of rubber with the magnitude of F along the ordinate axis and the absolute elongation Δl along the abscissa axis. Draw the conclusion about the relation between these quantities.

10. Calculate the modulus of elasticity E for rubber by the formula

$$E = \frac{l_0 F}{\Delta l S}$$

for two values of force (within the rectilinear part of the graph) and determine the average value of E.

11. Repeat the experiment with the rubber cord having a different cross section and determine again the modulus of elasticity of rubber for the same values of force.

4. Determination of EMF and the Internal Resistance of a Current Source

Instruments and materials: a current source (a battery or a cell for a flash-light), a rheostat, an ammeter, a voltmeter, and connecting wires.

HINTS. 1. Assemble a circuit connecting in series the battery, the rheostat, the ammeter and the key.

2. Connect the voltmeter to the terminals of the battery.

3. Measure the current in the circuit and the voltage in the inner part of the circuit.

4. Measure the resistance of the circuit with the help of the rheostat and then measure the voltage and the current once more.

5. Put the results into the table

Current I, A	Voltage U, V	EMF $\mathscr{E}$, V	Internal resistance of the source r, Ω

6. Using Ohm's law for a closed circuit $I = \mathscr{E}/(R + r)$ twice, determine the internal resistance and the emf of the current source using the data on the current and voltage. Find the resistance of the rheostat for each measurement.

7. Disconnect the circuit and measure the resistance of the battery. Compare the obtained result with the measured one.

5. Determination of Resistivity of a Conductor

Instruments and materials: an ammeter, a voltmeter, a measuring tape, an accumulator battery, vernier calipers, a Fechral wire, having the length of 65-70 cm and the diameter of 0.5 mm, with metallic tags, a key and connecting wires.

HINTS. 1. Measure the length of the wire with the measuring tape (the distance between the tags).

2. Measure the diameter of the wire and calculate its cross-sectional area.

3. Assemble the circuit by connecting in series the accumulator battery, the Fechral wire, the ammeter and the key.

4. Connect the voltmeter in parallel to the wire.

5. Having closed the key, measure the current in the circuit and the voltage across the wire ends.

6. Using Ohm's law, calculate the resistance of the wire.

7. Calculate the resistivity of fechral wire by the formula

$$\rho = \frac{RS}{l},$$

where R is the resistance of the wire, S is its cross-sectional area and l is its length.

6. Determination of Electrochemical Equivalent of Copper

The electrochemical equivalent can be determined with the help of Faraday's law of electrolysis. For this we must measure the mass of a substance deposited on the electrode, the current and the time during which it flows through the electrolyte.

Instruments and materials: a balance with weights, an ammeter, a clock, a hot plate, an accumulator, a rheostat, a switch, copper electrodes, a cylindrical vessel, copper sulphate solution, connecting wires, a test tube with sand, and a box for sand.

HINTS. 1. Put a copper electrode which will serve as a cathode on one pan of the balance and the box for sand on the other pan.

2. Thoroughly match the weights by carefully pouring sand from the test tube into the box.

3. Assemble the circuit by connecting in series the battery, the ammeter, the rheostat, the vessel with copper sulphate solution with copper electrodes immersed in it, and the switch.

4. Close the switch and note the time of the beginning of the experiment.

5. Maintaining a constant current (about 1 A) with the help of the rheostat, ensure that the pointer of the ammeter coincides with a scale division.

6. Carry out the electrolysis of the solution for 20 min.

7. Switch off the current, put out and rinse in water the copper electrode which has been used as a cathode, and dry it over a hot plate.

8. Put the electrode on the balance and match the weights again, but now with the help of weights whose mass is equal to the mass m of copper.

9. Tabulate the results.

Time t, s	Current J, A	Error in measurement of current ΔJ, A	Mass of copper m, kg

10. Using Faraday's law of electrolysis, determine the electrochemical equivalent for copper, $k = m/It$.

11. Determine the relative error in measurement of k from the formula $\varepsilon_k = \varepsilon_I + \varepsilon_m + \varepsilon_t$. Calculate $\varepsilon_I = \Delta I/I$. In calculating $\varepsilon_m = \Delta m/m$, we can assume that Δm does not exceed 20 mg. This error is determined by the sensitivity of the balance (it is equal to 10 mg for the electrode mass of about 20 g) and by the error of milligram weights, which in total is below 10 mg. The error $\varepsilon_t = \Delta t/t$ can be neglected.

12. Determine the absolute error in the measurement of the electrochemical equivalent for copper, $\Delta k = k\varepsilon_k$.

13. Represent the results of measurements in the form

$$k - \Delta k \leqslant k \leqslant k + \Delta k.$$

14. Using a handbook, determine the electrochemical equivalent for copper and check whether its value lies in the experimentally determined interval $|k - \Delta k, \ k + \Delta k|$.

15. Using the formula $k = (1/eN_A) \cdot (M/n)$, calculate e, viz. the magnitude of the elementary charge.

7. Observing the Effect Produced by a Magnetic Field on Current

Instruments and materials: an accumulator, a horse-shoe magnet, a holder, a switch, a wire coil and connecting wires.

HINTS. 1. Suspend the coil from the holder and connect it to the battery through the switch.

2. Bring the magnet close to the suspended coil, close the circuit and watch the motion of the coil.

3. Choose several typical versions of the relative arrangement of the coil and the magnet and draw them, indicating the direction of the magnetic field, the direction of current and the expected motion of the coil.

4. Verify experimentally the correctness of the assumptions about the nature and direction of motion of the coil.

8. Study of Electromagnetic Induction

Instruments and materials: a milliammeter, an accumulator, coils with cores, a horse-shoe magnet, a switch and connecting wires.

HINTS. 1. Connect the clamps of milliammeter with the terminals of the coil.

2. Attach the core to a pole of the horse-shoe magnet and insert it into the coil, watching the pointer of the milliammeter.

3. Repeat the observation, pulling the core out of the coil and then interchanging the poles of the magnet.

4. Draw a schematic diagram of the experiment and verify Lenz's rule in each case.

5. Place the second coil near the first one so that their axes coincide.

6. Insert the cores into two coils and connect the second coil to the battery through the key.

7. Connecting and disconnecting the circuit with the help of the key, watch the deflection of the pointer of the milliammeter.

8. Draw a schematic diagram of the experiment and verify Lenz's rule.

Answers to Exercises

Exercise 1. 1. Not more than 12 m². 2. 0.002 kg/mole, 0.004 kg/mole. 3. Twice. 4. $\simeq$ 0.056 mole. 5. $\simeq 1.88 \times 10^{23}$. 6. $\simeq 4.65 \times 10^{-26}$ kg. 7. $\simeq 8.5 \times 10^{28}$. 8. $\simeq$ 5.7×10^{-8} m³. 9. Decreases by a factor of 3. 10. 5×10^5 Pa. 11. 6×10^{-21} J. 12. 4.9×10^5 m²/s².

Exercise 2. 1. 2.76×10^{-23} J/K$_{new}$. 2. $\simeq 6 \times 10^{-21}$ J. 3. 3.14×10^4. 4. Air contains more molecules. 5. 5.3×10^{-26} kg. 6. $\simeq 0.5\%$.

Exercise 3. 1. 12 kPa. 2. 20. 4. 250 K. 6. $\simeq 0.024$ m³/mole. 8. $\simeq 5.8 \times 10^{-3}$ kg/mole. 9. $\simeq 0.15$ m³. 10. $\simeq 0.49$ kg/m³. 13. 1.7 times.

Exercise 4. 1. Increases 1.5 times. 2. 2.5×10^{-4} m³. 3. Decreases. 4. $\simeq 34.4$ J. 8. By $(m/M) R \Delta T$. 9. 1.25×10^6 J. 10. 20 J. 11. 2×10^5 J. 12. $\simeq 10$ K. 13. $\simeq 37°C$. 14. $0°C$. 15. 1500 K. 16. 20%; $\simeq 42\%$.

Exercise 5. 4. $\simeq 0.59$ kg/m³. 7. $\simeq 52\%$. 8. $\simeq 0.92$ kg. 9. $\simeq 0.21$ kg.

Exercise 6. 4. $\simeq 2.3 \times 10^{-5}$ N. 6. 800 kg/m³.

Exercise 7. 3. 8 mm². 4. $\simeq 1.9 \times 10^3$ N. 5. 3.53×10^5 Pa. 6. $\simeq 4200$ m. 7. 4×10^7 Pa, 2×10^{11} Pa. 8. $2/3$.

Exercise 8. 3. $\simeq 9.2 \times 10^{-8}$ N. 4. $\simeq 2.3 \times 10^{39}$. 5. $\simeq 2.3 \times 10^6$ N. 6. $\simeq 1.0 \times 10^{-6}$ N, attractive force; $\simeq 6.9 \times 10^{-7}$ N, repulsive force. 7. $\simeq 1.1 \times 10^{-6}$ N, towards the second charge.

Exercise 9. 1. $\simeq 1.5 \times 10^{-16}$ C; $\simeq 940$ excess electrons. 5. 1.6×10^{-19} J; -1.6×10^{-19} J. 6. $E = \sqrt{q_1^2 + q_1 q_2 + q_2^2}/4\pi\varepsilon_0 \varepsilon r^2$. 8. 4000 V/m. 10. -2.3×10^3 V.

Exercise 10. 1. $\simeq 0.1$ μF. 2. $\simeq 1.7 \times 10^{-5}$ C. 3. $\simeq 1.7 \times 10^7$ V/m. 4. $\simeq 5.5$ mm. 5. $U_1 = d_1 U/d\varepsilon = 100$ V. 7. Decreases by a factor of 3. 8. $\simeq 4.4 \times 10^{-4}$ J/m³. Exercise 11. 3. $\simeq 1.4$ mm²; $\simeq 15.8$ m. 4. Increases 25 times. 5. 500 Ω. 6. $\simeq 4.9 \times 10^{-4}$ m/s. 10. Reduced to half its value; doubled. 11. $\simeq 4.1$ kW, $\simeq 6.1$ kW. 12. $\simeq 97\%$, $\simeq 1670$ A.

Exercise 12. 2. 1200 A. 3. 3.7 V, 0.2 Ω. 4. $\simeq 0.33$ W. 6. 4.1 V, 22.4 Ω.

Exercise 13. 1. $\simeq 2.9 \times 10^4$ C. 2. $\simeq 3.3 \times 10^{-7}$ kg/C. 4. $\simeq 3 \times 10^{-5}$ m. 6. $\simeq 1.33 \times 10^7$ m/s, $\simeq 4.19 \times 10^7$ m/s.

Exercise 14. 3. 1.5×10^{-3} N·m. 4. 1.2×10^{-2} J. 6. $R = mv \sin\alpha/(|q|B)$.

Exercise 15. 6. In the second coil, a direct current will flow in a certain time. 8. 0.2 A. 9. 0.15 V. 10. 1.2 J.

TO THE READER

Substance	Permittivity
Air (at 0 °C and 760 mm Hg)	1.000594
Kerosene	2.1
Ebonite	2.7–2.9
Quartz	4.5
Glass	5–10
Ethyl alcohol	27
Water (pure)	81
Rochelle salt	10 000

Substance	Resistivity at 20 °C, $\Omega \cdot m$
Silver	1.6×10^{-8}
Copper	1.8×10^{-8}
Graphite	3×10^{-8}
Aqueous solution of NaCl (10%)	0.08
Silicon	10^{3}
Water (chemically pure)	10^{6}
Porcelain	10^{13}

Substance	Temperature of transition to superconducting state, K
Titanium	0.4
Uranium	0.8
Zinc	0.9
Aluminium	1.2
Tin	3.8
Mercury	4.1
Lead	7.2
Niobium nitrate	15.2

Substance	Temperature resistance coefficient, K^{-1}
Aluminium	3.8×10^{-3}
Iron	6.2×10^{-3}
Copper	4.3×10^{-3}
Mercury (liquid)	0.9×10^{-3}
Constantan (40% Ni, 60% Cu)	2×10^{-5}
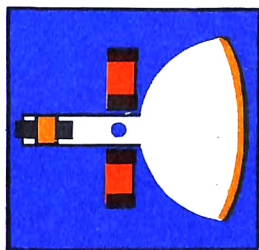
Substance	Temperature resistance coefficient, K^{-1}
Nichrome (20% Cr, 75% Ni, 5% Fe)	1×10^{-4}
Fecral (13% Cr, 4% Al, 1% Si, 0.7% Mn, the rest—Fe)	2×10^{-4}
Chromal (25% Cr, 5% Al, 70% Fe)	4×10^{-5}